

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:47:23 ; Search time 3410.29 Seconds
(without alignments)
17679.337 Million cell updates/sec

Title: US-09-989-981A-3
Perfect score: 2019
Sequence: 1 atggctgagaaaaccaaaga.....agtcaattcaagactggtga 2019

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*

28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	8					Description
	No.	Score	Query Match	Length	ID	
	1	2006	99.4	3623	11 AK004871	AK004871 Mus muscu
	2	1700	84.2	2417	11 AK050938	AK050938 Mus muscu
	3	714.2	35.4	849	12 BI330745	BI330745 602982409
	4	549	27.2	549	10 BF660076	BF660076 maa27c08.
	5	498	24.7	583	13 BY705076	BY705076 BY705076
	6	448.2	22.2	463	9 AA537862	AA537862 vj35a03.r
	7	446	22.1	510	10 BB610072	BB610072 BB610072
	8	442	21.9	511	9 AI157365	AI157365 ui45h01.y
	9	422.4	20.9	500	9 AI151811	AI151811 ui46c10.y
	10	398	19.7	398	9 AI597406	AI597406 vj35a03.y
	11	303.4	15.0	586	11 AK008188	AK008188 Mus muscu
	12	298.4	14.8	581	13 BY708144	BY708144 BY708144
c	13	280.8	13.9	781	14 CB502603	CB502603 ssalmge50
	14	275.8	13.7	435	9 AI574075	AI574075 uj67h11.y
	15	254	12.6	334	13 BX482362	BX482362 DKFZp686F
	16	250.4	12.4	640	14 CD739823	CD739823 4028769 1
	17	244	12.1	691	13 BX481838	BX481838 DKFZp686M
	18	179	8.9	916	29 CG323718	CG323718 OG0DQ45TH
c	19	177.4	8.8	709	29 CC695831	CC695831 OGULO23TV
	20	175.8	8.7	685	28 AZ339049	AZ339049 1M0070C22
c	21	172.6	8.5	762	29 CC659228	CC659228 OGUFF57TV
	22	172.6	8.5	821	28 BZ650554	BZ650554 OGCBAB89TC
	23	170.4	8.4	826	29 CG214497	CG214497 OG1BM08TV
c	24	168	8.3	891	29 CG368338	CG368338 OG3BP65TV
c	25	166.8	8.3	861	29 CG262933	CG262933 OG1DH53TV
	26	165.6	8.2	776	29 CG327545	CG327545 OGWFJ96TV
	27	164	8.1	345	14 CD730599	CD730599 4038931 1
	28	158	7.8	909	29 CG268466	CG268466 OG2BT15TH
	29	157	7.8	833	10 BF620684	BF620684 HVSMEc002
c	30	151.6	7.5	435	13 BX099922	BX099922 BX099922
c	31	151.6	7.5	912	29 CC604602	CC604602 OGUFQ75TH
	32	151.4	7.5	694	13 CA140253	CA140253 SCEZRT202
	33	150.6	7.5	303	10 BB870338	BB870338 BB870338
c	34	149.8	7.4	837	29 CC710016	CC710016 OGVBI31TH
	35	149.6	7.4	849	29 CG270361	CG270361 OGWFS70TH
	36	149.6	7.4	857	29 CG271003	CG271003 OG0EJ71TV
	37	149	7.4	447	12 BI145065	BI145065 602909138
	38	147.6	7.3	839	29 CG262656	CG262656 OG1AN46TH
	39	145.4	7.2	564	14 T84531	T84531 yd53b02.r1
	40	143.2	7.1	596	14 CA192267	CA192267 SCCCST3C0
	41	142.8	7.1	560	9 AU192604	AU192604 AU192604
	42	142	7.0	816	12 BG299523	BG299523 HVSMEa002
	43	141.6	7.0	936	10 BF162656	BF162656 601769307
	44	134.6	6.7	905	29 CG224339	CG224339 OGYAG39TV
c	45	134	6.6	490	28 BZ614705	BZ614705 ig46b01.g

ALIGNMENTS

RESULT 1

AK004871

LOCUS AK004871 3623 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300003C16 product:ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full insert sequence.

ACCESSION AK004871

VERSION AK004871.1 GI:12836380

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 3623)

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cdNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cdNA was primed with a primer [5' GAGAGAGAGAGCGCCGCAACTCGAGTTTTTTTTTTTTTTTTTTVN 3'], cdNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cdNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCCC 3']. cdNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

source Location/Qualifiers

1. .3623

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:1300003C16"

/db_xref="MGI:1896857"

/db_xref="taxon:10090"

/clone="1300003C16"

/sex="male"

/tissue_type="liver"

/clone_lib="RIKEN full-length enriched mouse cdNA library"

/dev_stage="adult"

CDS

69. .2090

/note="unnamed protein product; ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus] (SWISSPROT|Q9DBM0, evidence: FASTY, 92%ID, 96.7%length, match=1796)

putative"

/codon_start=1

/protein_id="BAB23630.1"

/db_xref="GI:12836381"


```

/translation="MAEKTKEETQLWNGTVLQDASQGLQDSLFSSES DNSLYFTYSGQ
SNTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSQML
AIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLP
NLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGE
RRRVSIGVQLLWNPGLILILDEPTSGLDSTAHNLVTLSRLAKGNRLVLISLHQPRSD
IFRLFDLVLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKE
REVATVEKAQSLAALFLEKVQGFDDFLWKA EAKELNTSTHTVSITLTQDTCGTAVEL
PGMIEQFSTLIRROI SNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLS FMDTAA
LLEMIGALIPFNVILDVVS KCHSERSMLYEELEDGLYTAGPYFFAKILGELPEHCAYV
IIYAMPIYWLTNLRPVPELFLHFLLVWLVVFCCRTMALAASAMLP TFHMS SFFCNAL
YNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSI
LGD TMISAMD LNSHPLYAIYLIVIGISYGF LFLYYLSLKLKQKSIQDW"

polyA_signal 3605. .3610
               /note="putative"
polyA_site    3623
               /note="putative"
ORIGIN

```

```

Query Match          99.4%;   Score 2006;   DB 11;   Length 3623;
Best Local Similarity 99.9%;   Pred. No. 0;
Matches 2019;   Conservative    0;   Mismatches    0;   Indels    3;   Gaps    1;

```

```

Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 128

Qy      61 TC---GGGCCTCCAGGACAGCTTGTCTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 TCGCAGGGCCTCCAGGACAGCTTGTCTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 188

Qy     118 TACAGTGGTCAGTCCAACACTCTGGAGGTGAGAGATCTCACCTACCAGGTGGACATCGCC 177
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 TACAGTGGTCAGTCCAACACTCTGGAGGTGAGAGATCTCACCTACCAGGTGGACATCGCC 248

Qy     178 TCTCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 237
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 TCTCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 308

Qy     238 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 297
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 368

Qy     298 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 357
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     369 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 428

Qy     358 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 417
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 488

Qy     418 ACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCC 477
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 ACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCC 548

Qy     478 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTC 537
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTC 608

```

Qy	538	TCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAG	597
Db	609	TCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAG	668
Qy	598	TGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGA	657
Db	669	TGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGA	728
Qy	658	CGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCC	717
Db	729	CGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCC	788
Qy	718	ACTTCTGGCCTCGACAGCTTCACAGCCCACAACTCTGGTGACAACCTTGTCCCGCCTGGCC	777
Db	789	ACTTCTGGCCTCGACAGCTTCACAGCCCACAACTCTGGTGACAACCTTGTCCCGCCTGGCC	848
Qy	778	AAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTA	837
Db	849	AAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTA	908
Qy	838	TTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAA	897
Db	909	TTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAA	968
Qy	898	ATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGAC	957
Db	969	ATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGAC	1028
Qy	958	TTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTG	1017
Db	1029	TTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTG	1088
Qy	1018	GAGAAGGCACAGTCTCTTGCAGCCCTGTTCCTAGAAAAAGTACAAGGCTTTGATGACTTT	1077
Db	1089	GAGAAGGCACAGTCTCTTGCAGCCCTGTTCCTAGAAAAAGTACAAGGCTTTGATGACTTT	1148
Qy	1078	CTGTGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTGACCCTC	1137
Db	1149	CTGTGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTGACCCTC	1208
Qy	1138	ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCC	1197
Db	1209	ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCC	1268
Qy	1198	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTTCAT	1257
Db	1269	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTTCAT	1328
Qy	1258	GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1317
Db	1329	GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1388
Qy	1318	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1377
Db	1389	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1448

Qy	1378	TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACCTCGGAGAGGTCAATGCTGTACTAT	1437
Db	1449	TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACCTCGGAGAGGTCAATGCTGTACTAT	1508
Qy	1438	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA	1497
Db	1509	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA	1568
Qy	1498	TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1557
Db	1569	TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1628
Qy	1558	CTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC	1617
Db	1629	CTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC	1688
Qy	1618	TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCC	1677
Db	1689	TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCC	1748
Qy	1678	TTCTTCTGCAATGCCCTCTACAACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG	1737
Db	1749	TTCTTCTGCAATGCCCTCTACAACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG	1808
Qy	1738	GACAACTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTC	1797
Db	1809	GACAACTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTC	1868
Qy	1798	TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC	1857
Db	1869	TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC	1928
Qy	1858	ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC	1917
Db	1929	ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC	1988
Qy	1918	TATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA	1977
Db	1989	TATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA	2048
Qy	1978	TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019
Db	2049	TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2090

RESULT 2

AK050938

LOCUS AK050938 2417 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030040P06 product:ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full insert sequence.

ACCESSION AK050938

VERSION AK050938.1 GI:26094211

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 2417)
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/.

FEATURES Location/Qualifiers
source 1. .2417
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D030040P06"
/db_xref="MGI:2418860"
/db_xref="taxon:10090"
/clone="D030040P06"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9 days embryo"
misc_feature 1. .2417
/note="ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus] (SWISSPROT|Q9DBM0, evidence: FASTY, 92%ID, 96.7%length, match=1796)"

ORIGIN

Query Match 84.2%; Score 1700; DB 11; Length 2417;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1700; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	320	CAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGA	379
Db	184	CAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGA	243
Qy	380	TGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGT	439
Db	244	TGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGT	303
Qy	440	GCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCC	499
Db	304	GCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCC	363
Qy	500	TGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCCAGCGTGACAAAC	559
Db	364	TGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCCAGCGTGACAAAC	423
Qy	560	GGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCA	619
Db	424	GGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCA	483
Qy	620	ACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGC	679
Db	484	ACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGC	543

Qy	680	TCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCA	739
Db	544	TCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCA	603
Qy	740	CAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCA	799
Db	604	CAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCA	663
Qy	800	TCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGA	859
Db	664	TCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGA	723
Qy	860	CATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCA	919
Db	724	CATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCA	783
Qy	920	TTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCA	979
Db	784	TTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCA	843
Qy	980	TCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCCTGGAGAAGGCACAGTCTCTTGCAG	1039
Db	844	TCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCCTGGAGAAGGCACAGTCTCTTGCAG	903
Qy	1040	CCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTGTGGAAAGCTGAGGCAAAGG	1099
Db	904	CCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTGTGGAAAGCTGAGGCAAAGG	963
Qy	1100	AACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACACAGGACACTGACTGTGGGA	1159
Db	964	AACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACACAGGACACTGACTGTGGGA	1023
Qy	1160	CTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTTCCACCCTGATCCGTCGTCAGATTT	1219
Db	1024	CTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTTCCACCCTGATCCGTCGTCAGATTT	1083
Qy	1220	CCAATGACTTCCGGGACCTGCCCACGCTGCTCATTATGGGTTCGGAAGCCTGCCTGATGT	1279
Db	1084	CCAATGACTTCCGGGACCTGCCCACGCTGCTCATTATGGGTTCGGAAGCCTGCCTGATGT	1143
Qy	1280	CCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCCAAGCAGCTCTCCTTCATGGACA	1339
Db	1144	CCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCCAAGCAGCTCTCCTTCATGGACA	1203
Qy	1340	CAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTCAATGTCATCCTGGATGTCG	1399
Db	1204	CAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTCAATGTCATCCTGGATGTCG	1263
Qy	1400	TCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGTACA	1459
Db	1264	TCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGTACA	1323
Qy	1460	CTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACG	1519
Db	1324	CTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACG	1383

Qy 1520 TCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTGCGGCCCGTGCCTGAGCTCT 1579
 |||
 Db 1384 TCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTGCGGCCCGTGCCTGAGCTCT 1443
 |||
 Qy 1580 TCCTTCTACACTTCCTGCTCGTGTGGTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGG 1639
 |||
 Db 1444 TCCTTCTACACTTCCTGCTCGTGTGGTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGG 1503
 |||
 Qy 1640 CTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACA 1699
 |||
 Db 1504 CTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACA 1563
 |||
 Qy 1700 ACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTG 1759
 |||
 Db 1564 ACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTG 1623
 |||
 Qy 1760 CATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGGTCTTCTCGGGGCTGATGCAGATTCAAT 1819
 |||
 Db 1624 CATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGGTCTTCTCGGGGCTGATGCAGATTCAAT 1683
 |||
 Qy 1820 TTAATGGACACCTTTACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACA 1879
 |||
 Db 1684 TTAATGGACACCTTTACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACA 1743
 |||
 Qy 1880 CGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCA 1939
 |||
 Db 1744 CGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCA 1803
 |||
 Qy 1940 TCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGA 1999
 |||
 Db 1804 TCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGA 1863
 |||
 Qy 2000 AGTCAATTCAAGACTGGTGA 2019
 |||
 Db 1864 AGTCAATTCAAGACTGGTGA 1883
 |||

RESULT 3

BI330745

LOCUS BI330745 849 bp mRNA linear EST 30-JUL-2001

DEFINITION 602982409F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5135115 5', mRNA sequence.

ACCESSION BI330745

VERSION BI330745.1 GI:15015402

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 849)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11332 row: a column: 04
 High quality sequence stop: 758.

FEATURES Location/Qualifiers
 source 1. .849
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5135115"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Li9"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 35.4%; Score 714.2; DB 12; Length 849;
 Best Local Similarity 95.9%; Pred. No. 1.9e-163;
 Matches 799; Conservative 0; Mismatches 23; Indels 11; Gaps 6;

```

Qy      891 GCAGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCC 950
         |||
Db      1   GCAGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCC 60

Qy      951 TGCGGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGC 1010
         |||
Db      61  TGCGGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGC 120

Qy      1011 CACCGTGGAGAAGGCACAGTCTCTTGCAGCCCTGTTCCCTAGAAAAAGTACAAGGCTTTGA 1070
         |||
Db      121 CACCGTGGAGAAGGCACAGTCTCTTGCAGCCCTGTTCCCTAGAAAAAGTACAAGGCTTTGA 180

Qy      1071 TGACTTTCTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCT 1130
         |||
Db      181 TGACTTTCTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCT 240

Qy      1131 GACCCTCACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCA 1190
         |||
Db      241 GACCCTCACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCA 300

Qy      1191 GTTTTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCT 1250
         |||
Db      301 GTTTTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCT 360

Qy      1251 CATTTCATGGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCA 1310
         |||
Db      361 CATTTCATGGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCA 420

Qy      1311 TGGGGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCT 1370
         |||
Db      421 TGGGGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCT 480
  
```


Qy 1371 CATTCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACCTCGGAGAGGTCAATGCT 1430
 |||
 Db 481 CATTCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACCTCGGAGAGGTCAATGCT 540
 Qy 1431 GTACTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCT 1490
 |||
 Db 541 GTACTATGAGCTGGAAGACGGGCTGT--ACTGCTGGTCCTTATTTCTTTGCCAAGATCCT 598
 Qy 1491 AGGAGAATTGCCGGAGCAC-TGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGC 1549
 |||
 Db 599 AGGAGAATTGCCGGAGCACTTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGC 658
 Qy 1550 TGACAAACCTGCGGCCCCGTGCCTGAGCTCTTCCTTCTAC--ACTTCCTGCTCGTGTGGTT 1607
 |||
 Db 659 TGACAAACCTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACCACTTTCCTGCTCGTGTGGTA 718
 Qy 1608 GGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCA 1667
 ||
 Db 719 GGAGGTCTTCTGCTGCAGGACATGGCCTTGG---TGCTCTGCCATGCTG-CCAACTTCCA 774
 Qy 1668 CATGTCCTCCTTCTTCTGCAATGCCCTCTACAACTCCTTCTACCTTACTGCCG 1720
 |||
 Db 775 CATGTCCTCCTTCTTCTGCA--TGCCTCTTAGAATCCTTCTACCTTATGGCGG 825

RESULT 4

BF660076

LOCUS BF660076 549 bp mRNA linear EST 20-DEC-2000

DEFINITION maa27c08.y1 NCI_CGAP_Li10 Mus musculus cDNA clone IMAGE:3812342 5' similar to TR:Q9VQN4 Q9VQN4 CG9664 PROTEIN. ;, mRNA sequence.

ACCESSION BF660076

VERSION BF660076.1 GI:11925210

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 549)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Other_ESTs: maa27c08.x1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

MGI:1454454

Seq primer: -40RP from Gibco

High quality sequence stop: 435.

```

FEATURES                      Location/Qualifiers
    source                    1. .549
                              /organism="Mus musculus"
                              /mol_type="mRNA"
                              /db_xref="taxon:10090"
                              /clone="IMAGE:3812342"
                              /sex="female"
                              /dev_stage="10 weeks"
                              /lab_host="DH10B (T1 phage-resistant)"
                              /clone_lib="NCI CGAP_Li10"
                              /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                              Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                              Average insert size 1.6 kb. Library constructed by Life
                              Technologies."

```

ORIGIN

```

Query Match          27.2%; Score 549; DB 10; Length 549;
Best Local Similarity 100.0%; Pred. No. 3.9e-123;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1462 GCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTC 1521
          |||
Db      1    GCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTC 60

Qy      1522 ATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTGCGGCCCCGTGCCTGAGCTCTTC 1581
          |||
Db      61   ATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTGCGGCCCCGTGCCTGAGCTCTTC 120

Qy      1582 CTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCT 1641
          |||
Db      121  CTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCT 180

Qy      1642 GCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAAC 1701
          |||
Db      181  GCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAAC 240

Qy      1702 TCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCA 1761
          |||
Db      241  TCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCA 300

Qy      1762 TGGATCTCCAAGCTGTCTGTTCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTT 1821
          |||
Db      301  TGGATCTCCAAGCTGTCTGTTCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTT 360

Qy      1822 AATGGACACCTTTACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACG 1881
          |||
Db      361  AATGGACACCTTTACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACG 420

Qy      1882 ATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATC 1941
          |||
Db      421  ATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATC 480

Qy      1942 GGCATCAGCTACGGCTTCCTGTTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAG 2001
          |||
Db      481  GGCATCAGCTACGGCTTCCTGTTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAG 540

Qy      2002 TCAATTCAA 2010

```

|||||||
Db 541 TCAATTCAA 549

RESULT 5
BY705076
LOCUS BY705076 583 bp mRNA linear EST 16-DEC-2002
DEFINITION BY705076 RIKEN full-length enriched, adult male liver Mus musculus
cDNA clone 1300003C16 5', mRNA sequence.
ACCESSION BY705076
VERSION BY705076.1 GI:27116215
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 583)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasaki,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,

URL:<http://genome.gsc.riken.go.jp/>

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome*. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
 source Location/Qualifiers
 1. .583
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1300003C16"
 /sex="male"
 /tissue_type="liver"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult male liver"

ORIGIN

Query Match	24.7%;	Score 498;	DB 13;	Length 583;
Best Local Similarity	99.4%;	Pred. No. 1.2e-110;		
Matches	511;	Conservative	0;	Mismatches 0; Indels 3; Gaps 1;
Qy	1	ATGGCTGAGAAAACCAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT	60	
Db	69	ATGGCTGAGAAAACCAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT	128	
Qy	61	TC---GGGCCTCCAGGACAGCTTGTCTCTCTCGGAAAGTGACAACAGTCTGTACTTCACC	117	
Db	129	TCGCAGGGCCTCCAGGACAGCTTGTCTCTCTCGGAAAGTGACAACAGTCTGTACTTCACC	188	
Qy	118	TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC	177	
Db	189	TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC	248	
Qy	178	TCTCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC	237	

```

      |||
Db      249 TCTCAGGTGCCTTGGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 308
      |||
Qy      238 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 297
      |||
Db      309 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 368
      |||
Qy      298 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 357
      |||
Db      369 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 428
      |||
Qy      358 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 417
      |||
Db      429 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 488
      |||
Qy      418 ACGCCTCAGCTGGTGAGGAAGTGCCTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCC 477
      |||
Db      489 ACGCCTCAGCTGGTGAGGAAGTGCCTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCC 548
      |||
Qy      478 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTG 511
      |||
Db      549 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTG 582

```

RESULT 6

AA537862

LOCUS AA537862 463 bp mRNA linear EST 29-JUL-1997

DEFINITION vj35a03.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA clone IMAGE:930988 5', mRNA sequence.

ACCESSION AA537862

VERSION AA537862.1 GI:2283855

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 463)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:535908

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 393.

FEATURES

Location/Qualifiers

```

source      1. .463
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:930988"
            /tissue_type="diaphragm"
            /dev_stage="adult"
            /lab_host="SOLR (kanamycin resistant)"
            /clone_lib="Stratagene mouse diaphragm (#937303)"
            /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
            EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
            prepared from diaphragm muscle. Primer: Oligo dT. Average
            insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
            sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
            CTCGAGTTTTTTTTTTTTTTTTTTT 3'"

```

ORIGIN

```

Query Match      22.2%; Score 448.2; DB 9; Length 463;
Best Local Similarity 98.3%; Pred. No. 1.7e-98;
Matches 453; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

Qy      1018 GAGAAGGCACAGTCTCTTGAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTT 1077
          |||
Db        1 GAGAAGGCACAGTCTCTTGAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTT 60

Qy      1078 CTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCCTC 1137
          |||
Db        61 CTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCCTC 120

Qy      1138 ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTC 1197
          |||
Db       121 ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTC 180

Qy      1198 ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT 1257
          |||
Db       181 ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT 240

Qy      1258 GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCC 1317
          |||
Db       241 GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCA 300

Qy      1318 AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT 1377
          |||
Db       301 GAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT 360

Qy      1378 TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT 1437
          |||
Db       361 TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT 420

Qy      1438 GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTT 1478
          |||
Db       421 GAGCTGGAAGACGGGCTGTACACTGCCAATACATATTTCTT 461

```

RESULT 7

BB610072

LOCUS BB610072 510 bp mRNA linear EST 26-OCT-2001

DEFINITION BB610072 RIKEN full-length enriched, adult male liver *Mus musculus* cDNA clone 1300007N20 5', mRNA sequence.
 ACCESSION BB610072
 VERSION BB610072.1 GI:16451685
 KEYWORDS EST.
 SOURCE *Mus musculus* (house mouse)
 ORGANISM *Mus musculus*
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 510)
 AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
 TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,H., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* . 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* . 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* . 11 (2), 281-289 (2001)
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
 e mouse tissues.
 FEATURES Location/Qualifiers
 source 1. .510
 /organism="Mus musculus"
 /mol_type="mRNA"

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1300007N20"
/sex="male"
/tissue_type="liver"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male liver"

ORIGIN

```
Query Match          22.1%;  Score 446;  DB 10;  Length 510;
Best Local Similarity 100.0%;  Pred. No. 6.1e-98;
Matches 446;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 123

Qy      61 TCGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 TCGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 183

Qy     121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 243

Qy     181 CAGGTGCCTTGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 CAGGTGCCTTGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 303

Qy     241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 363

Qy     301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 423

Qy     361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 483

Qy     421 CCTCAGCTGGTGAGGAAGTGC GTTGC 446
        ||||||||||||||||||
Db     484 CCTCAGCTGGTGAGGAAGTGC GTTGC 509
```

RESULT 8
AI157365

LOCUS AI157365 511 bp mRNA linear EST 30-SEP-1998
DEFINITION ui45h01.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1885393 5', mRNA sequence.
ACCESSION AI157365
VERSION AI157365.1 GI:3685834
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 511)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:969717
 Seq primer: custom primer used
 High quality sequence stop: 480.
 FEATURES Location/Qualifiers
 source 1. .511
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1885393"
 /dev_stage="embryo, 14 dpc"
 /lab_host="DH10B"
 /clone_lib="Sugano mouse embryo mewa"
 /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
 Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
 with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTTT];
 double-stranded cDNA was ligated to a DraIII adaptor
 [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
 sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
 CACCATGTG). XhoI should be used to isolate the cDNA
 insert. Size selection was performed to exclude fragments
 <1.5kb. Library constructed by Dr. Sumio Sugano
 (University of Tokyo Institute of Medical Science).
 Custom primers for sequencing: 5' end primer
 CTTCTGCTCTAAAAGCTGCG and 3' end primer
 CGACCTGCAGCTCGAGCACA."

ORIGIN

Query Match 21.9%; Score 442; DB 9; Length 511;
 Best Local Similarity 99.3%; Pred. No. 5.9e-97;
 Matches 455; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
 |||
 Db 54 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 113
 Qy 61 TC---GGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 117
 ||
 Db 114 TCGCAGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 173

Qy 118 TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC 177
 |||
 Db 174 TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC 233
 Qy 178 TCTCAGGTGCCTTGGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 237
 |||
 Db 234 TCTCAGGTGCCTTGGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 293
 Qy 238 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 297
 |||
 Db 294 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 353
 Qy 298 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 357
 |||
 Db 354 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 413
 Qy 358 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 417
 |||
 Db 414 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 473
 Qy 418 ACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCG 455
 |||
 Db 474 ACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCG 511

RESULT 9

AI151811

LOCUS AI151811 500 bp mRNA linear EST 30-SEP-1998

DEFINITION ui46c10.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
 IMAGE:1885458 5', mRNA sequence.

ACCESSION AI151811

VERSION AI151811.1 GI:3680280

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 500)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:969782

Seq primer: custom primer used

High quality sequence stop: 499.

FEATURES

source

Location/Qualifiers

1. .500

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="IMAGE:1885458"

/dev_stage="embryo, 14 dpc"

/lab_host="DH10B"

/clone_lib="Sugano mouse embryo mewa"

/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

ORIGIN

Query Match 20.9%; Score 422.4; DB 9; Length 500;

Best Local Similarity 99.1%; Pred. No. 3.6e-92;

Matches 436; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

Qy 1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60

Db 61 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 120

Qy 61 TC---GGGCCTCCAGGACAGCTTGTCTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 117

Db 121 TCGCAGGGCCTCCAGGACAGCTTGTCTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 180

Qy 118 TACAGTGGTCAGTCCAACACTCTGGAGGTGAGAGATCTCACCTACCAGGTGGACATCGCC 177

Db 181 TACAGTGGTCAGTCCAACACTCTGGAGGTGAGAGATCTCACCTACCAGGTGGACATCGCC 240

Qy 178 TCTCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 237

Db 241 TCTCAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 300

Qy 238 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 297

Db 301 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 360

Qy 298 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 357

Db 361 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 420

Qy 358 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 417

Db 421 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 480

Qy 418 ACGCCTCAGCTGGTGAGGAA 437
 |||||
 Db 481 ACGCCTCAGCTGGTGAAGAA 500

RESULT 10

AI597406

LOCUS AI597406 398 bp mRNA linear EST 21-APR-1999

DEFINITION vj35a03.y1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA clone IMAGE:930988 5', mRNA sequence.

ACCESSION AI597406

VERSION AI597406.1 GI:4606454

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 398)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:535908

This read is a RESEQUENCE of a previously sequenced mouse clone

This read has been verified (found to hit its original self in the correct orientation)

Seq primer: -40RP from Gibco

High quality sequence stop: 374.

FEATURES

source

Location/Qualifiers

1. .398

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:930988"

/tissue_type="diaphragm"

/dev_stage="adult"

/lab_host="SOLR (kanamycin resistant)"

/clone_lib="Stratagene mouse diaphragm (#937303)"

/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA

prepared from diaphragm muscle. Primer: Oligo dT. Average

insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor

sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 19.7%; Score 398; DB 9; Length 398;
 Best Local Similarity 100.0%; Pred. No. 2.9e-86;
 Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1018 GAGAAGGCACAGTCTCTTGCAGCCCTGTTTCCTAGAAAAAGTACAAGGCTTTGATGACTTT 1077
          |||
Db      1    GAGAAGGCACAGTCTCTTGCAGCCCTGTTTCCTAGAAAAAGTACAAGGCTTTGATGACTTT 60

Qy      1078 CTGTGGAAAGCTGAGGCAAAGGAAGTCAACACAAGCACCCACACAGTCAGCCTGACCCTC 1137
          |||
Db      61    CTGTGGAAAGCTGAGGCAAAGGAAGTCAACACAAGCACCCACACAGTCAGCCTGACCCTC 120

Qy      1138 ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCC 1197
          |||
Db      121   ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCC 180

Qy      1198 ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCAT 1257
          |||
Db      181   ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCAT 240

Qy      1258 GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC 1317
          |||
Db      241   GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC 300

Qy      1318 AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT 1377
          |||
Db      301   AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT 360

Qy      1378 TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTC 1415
          |||
Db      361   TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTC 398
  
```

RESULT 11

AK008188

LOCUS AK008188 586 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010011G12 product:ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full insert sequence.

ACCESSION AK008188

VERSION AK008188.1 GI:12842221

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE

2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 586)
 AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
 Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
 Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
 Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
 Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
 Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
 Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
 Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
 Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
 Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission
 JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer

[5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```

FEATURES
    source          Location/Qualifiers
                    1. .586
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6J"
                    /db_xref="FANTOM_DB:2010011G12"
                    /db_xref="MGI:1897592"
                    /db_xref="taxon:10090"
                    /clone="2010011G12"
                    /sex="male"
                    /tissue_type="small intestine"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="adult"
    CDS             <1. .306
                    /note="unnamed protein product; ATP-BINDING CASSETTE,
                    SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus]
                    (SWISSPROT|Q9DBM0, evidence: FASTY, 92%ID, 96.7%length,
                    match=1796)
                    putative"
                    /codon_start=1
                    /protein_id="BAC25204.1"
                    /db_xref="GI:26359608"
                    /translation="PAGFMINLDNLWIVPAWISKLSFLRWCFSGVMQIQFNHLYTTQ
                    IGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISYGFLFLYYLSLKLKIQKSIQDW"

```

ORIGIN

Query Match 15.0%; Score 303.4; DB 11; Length 586;
 Best Local Similarity 99.7%; Pred. No. 5.5e-63;
 Matches 304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1715 CTGCCGGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGC 1774
          |||
Db      2   CTGCCGGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGC 61

Qy      1775 TGTCGTTCCCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTT 1834
          |||
Db      62   TGTCGTTCCCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTT 121

Qy      1835 ACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCA 1894
          |||
Db      122 ACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCA 181

Qy      1895 TGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATCGGCATCAGCTACG 1954
          |||
Db      182 TGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATCGGCATCAGCTACG 241

Qy      1955 GCTTCCTGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACT 2014
          |||

```

Db 242 GCTTCCTGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACT 301

Qy 2015 GGTGA 2019
 |||||

Db 302 GGTGA 306

RESULT 12

BY708144

LOCUS BY708144 581 bp mRNA linear EST 16-DEC-2002
 DEFINITION BY708144 RIKEN full-length enriched, adult male small intestine Mus musculus cDNA clone 2010011G12 5', mRNA sequence.

ACCESSION BY708144

VERSION BY708144.1 GI:27119328

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 581)

AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
 Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
 Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1. .581

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="2010011G12"

/sex="male"

/tissue_type="small intestine"

/dev_stage="adult"

/lab_host="SOLR"

/clone_lib="RIKEN full-length enriched, adult male small intestine"

/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI."

ORIGIN

Query Match 14.8%; Score 298.4; DB 13; Length 581;
 Best Local Similarity 99.7%; Pred. No. 9.1e-62;
 Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1720 GGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCTG 1779
          |||
Db      1    GGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCTG 60

Qy      1780 TTCCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACC 1839
          |||
Db      61   TTCCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACC 120

Qy      1840 ACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGAC 1899
          |||
Db      121  ACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGAC 180

Qy      1900 CTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTC 1959
          |||
Db      181  CTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTC 240

Qy      1960 CTGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
          |||
Db      241  CTGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 300
  
```

RESULT 13

CB502603/c

LOCUS CB502603 781 bp mRNA linear EST 16-MAY-2003

DEFINITION ssalmge503002 gut *Salmo salar* cDNA, mRNA sequence.

ACCESSION CB502603

VERSION CB502603.1 GI:29313829

KEYWORDS EST.

SOURCE *Salmo salar* (Atlantic salmon)

ORGANISM *Salmo salar*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; *Salmo*.

REFERENCE 1 (bases 1 to 781)

AUTHORS GRASP Consortium, Davidson, W.S., Koop, B.F. and
<http://web.uvic.ca/cbr/grasp>.

TITLE A survey of *Salmo salar* transcripts from high complexity cDNA
 libraries

JOURNAL Unpublished (2002)

COMMENT Contact: Koop BF
 Centre for Biomedical Research
 University of Victoria
 PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
 Tel: 250 472 4067
 Fax: 250 472 4075

Email: bkoop@uvic.ca

Genome Sciences Centre, BC Cancer Agency cDNA preparation,
 sequencing and bioinformatics: Y Butterfield, R Kirkpatrick, J
 Asano, N Girn, R Guin, D Lee, S Lee, T Olson, P Pandoh, A Prahbu, D
 Smailus, L Spence, J Stott, S Taylor, G Yang, J Schein, S Jones and
 M Marra.

POLYA=Yes.

FEATURES
source Location/Qualifiers
1. .781
/organism="Salmo salar"
/mol_type="mRNA"
/strain="McConnell"
/db_xref="taxon:8030"
/clone_lib="gut"
/note="Vector: pBlueScriptIISK+; Library Creator: Matthew L Rise ; Atlantic salmon tissue contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton, B.C.), Rachel Roper (University of Victoria)"

ORIGIN

Query Match 13.9%; Score 280.8; DB 14; Length 781;
Best Local Similarity 66.2%; Pred. No. 2.2e-57;
Matches 405; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```
Qy      1408 TGTCACCTCGGAGAGGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGTACACTGCTGGT 1467
          ||||| | |||| | ||| ||||| ||||| ||||| ||||| ||||| | |
Db      769 TGTCACACAGAGAGAGCTATGTTGTACCATGAGCTGGAGGACGGCATGTATAACGTCACA 710

Qy      1468 CCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTCATCATC 1527
          | || ||||| ||||| |||| | || | ||| ||||| |||| | ||
Db      709 TCCTACTTCTTTGCCAAGGTCTGGGGGAGCTTCCAGAGCACTGTGTGTTACGTTGGTC 650

Qy      1528 TACGCGATGCCCATCTACTGGCTGACAAACCTGCGGCCCGTGCCTGAGCTCTTCTTCTA 1587
          |||| | ||||| ||||| |||| | || | ||| ||||| |||| | ||
Db      649 TACGGCCTACCCATCTACTGGCTGGCTGGCCTGAACCAGGCCCGGACCGCTTCTTGCTC 590

Qy      1588 CACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCT 1647
          ||||| ||||| | |||| | |||| | || | ||||| ||| || ||
Db      589 AACTTCCTGCTGGTGTGGCTCATGGTGTACTGCAGCCGCAGCATGGCTCTGTTGTGGCT 530

Qy      1648 GCCATGCTGCCCACCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAACCTCCTTC 1707
          || | ||||| | || | || ||||| | |||| | || | || ||||
Db      529 GCAGCCTTACCCACCCTGCAGACATCAGCCTTCATGGGCAATTCTCTGTTCACTGTGTTTC 470

Qy      1708 TACCTTACTGCCGGCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATC 1767
          ||||| ||||| | || | |||| | |||| | ||| | |||| ||
Db      469 TACCTTACTGGAGGCTTCGTCATTAGCCTGGAGAACATGTGGTTCGTGGCGTCCTGGTTC 410

Qy      1768 TCCAAGCTGTCGTTCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTAAATGGA 1827
          ||| | || ||| | || |||| |||| |||| | |||| | || | |||
Db      409 TCCCATGCCTCCTTCATGCGCTGGGGCTTTGAGGGCATGCTGCAGGTCCAGTTCAGGGGA 350

Qy      1828 CACCTTTACACCACACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGATGATC 1887
          | | ||| || ||||| ||||| || |||| || || || || ||
Db      349 CTCAAGTACCCCGTCTCCATCGGCAACTTCACCCTCAACATCGATGGCATAACATGTGGTG 290

Qy      1888 AGTGCCATGGACCTGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATCGGCATC 1947
          || ||||| ||||| | || ||||| | ||||| || ||||| ||
Db      289 GAAGCTATGGATATGAACCACTACCCTCTCTACTCCTGCTACCTGGTTCTCATCGCTGTC 230

Qy      1948 AGCTACGGCTTCCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 2007
          ||||| || | || ||||| ||||| | || ||||| ||||| ||
```

Db 229 GTAGTGGGCTTCATGCTGCTCTACTACCTATCACTCAAATTCATCAAGCAGAAGTCCAGC 170

Qy 2008 CAAGACTGGTGA 2019
 || |||||

Db 169 CAGGACTGGTGA 158

RESULT 14

AI574075

LOCUS AI574075 435 bp mRNA linear EST 29-MAR-1999

DEFINITION uj67h11.y1 Sugano mouse liver mlia Mus musculus cDNA clone
 IMAGE:1925061 5', mRNA sequence.

ACCESSION AI574075

VERSION AI574075.1 GI:4537449

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 435)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
 Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:981353

Seq primer: custom primer used

High quality sequence stop: 432.

FEATURES

source

Location/Qualifiers

1. .435

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL"

/db_xref="taxon:10090"

/clone="IMAGE:1925061"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="Sugano mouse liver mlia"

/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
 (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
 was primed with an oligo(dT) primer

[ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested
 and cloned into distinct DraIII sites of the pME18S-FL3
 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
 be used to isolate the cDNA insert. Size selection was

performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

ORIGIN

```

Query Match          13.7%;  Score 275.8;  DB 9;  Length 435;
Best Local Similarity 99.3%;  Pred. No. 2.6e-56;
Matches 277;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

Qy      1  ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2  ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 61

Qy     61  TCGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62  TCGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 121

Qy    121  AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122  AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 181

Qy    181  CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182  CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 241

Qy    241  CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTC 279
        ||||||||||||||||||||||||||||||||||||
Db    242  CAAGACTCCTGTGAGCTGGGCATCCGAAATCTTACTTTC 280

```

RESULT 15

BX482362

LOCUS BX482362 334 bp mRNA linear EST 04-SEP-2003

DEFINITION DKFZp686F02230_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686F02230 5', mRNA sequence.

ACCESSION BX482362

VERSION BX482362.1 GI:31942182

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 334)

AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.

TITLE EST (Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., et al.)

JOURNAL Unpublished (2003)

COMMENT Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by BMFZ (Biomedical Research Center at the Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA sequencing

consortium of the German Genome Project. No s1 sequence available.
 This clone (DKFZp686F02230) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

```
FEATURES                      Location/Qualifiers
    source                    1. .334
                              /organism="Homo sapiens"
                              /mol_type="mRNA"
                              /db_xref="taxon:9606"
                              /clone="DKFZp686F02230"
                              /dev_stage="adult"
                              /lab_host="DH10B"
                              /clone_lib="686 (synonym: hlcc3)"
                              /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                              cDNA-collection"
```

ORIGIN

Query Match 12.6%; Score 254; DB 13; Length 334;
 Best Local Similarity 85.0%; Pred. No. 4.7e-51;
 Matches 284; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy      1189 CAGTTTTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTG 1248
          ||||| | | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1   CAGTTTACGACGCTGAGCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCCACCCTC 60

Qy      1249 CTCATTCATGGGTCGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGC 1308
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61 CTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGC 120

Qy      1309 CATGGGGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCG 1368
          ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 CATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCT 180

Qy      1369 CTCATTCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATG 1428
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      181 CTCATCCCTTTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATG 240

Qy      1429 CTGTACTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATC 1488
          || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241 CTTTACTATGAAGTGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATC 300

Qy      1489 CTAGGAGAATTGCCGGAGCACTGTGCCTACGTCA 1522
          || || || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      301 CTCGGGGAGCTTCCGGAGCACTGTGCCTACATCA 334
```

Search completed: February 26, 2004, 09:39:24
 Job time : 3419.29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:40:23 ; Search time 5185.97 Seconds
(without alignments)
16874.299 Million cell updates/sec

Title: US-09-989-981A-3
Perfect score: 2019
Sequence: 1 atggctgagaaaaaccaaaga.....agtcaattcaagactggtga 2019

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*

```

28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		%							
Result		Query							
No.	Score	Match	Length	DB	ID			Description	
	1	2019	100.0	2019	6	AX685731		AX685731 Sequence	
	2	2019	100.0	2284	10	AY196216		AY196216 Mus muscu	
	3	2006	99.4	3674	10	AF324495		AF324495 Mus muscu	
	4	1999.8	99.0	2285	10	AY196215		AY196215 Mus muscu	
	5	1727.8	85.6	4829	10	AF351785		AF351785 Rattus no	
	6	1430	70.8	2669	6	AX685735		AX685735 Sequence	
	7	1428.4	70.7	2022	9	AF320294		AF320294 Homo sapi	
	8	1428.4	70.7	2679	9	AF324494		AF324494 Homo sapi	
	9	743.8	36.8	3239	6	AX478099		AX478099 Sequence	
c	10	302.2	15.0	204584	10	AC122243		AC122243 Mus muscu	
	11	275.8	13.7	1387	10	F351799S06		AF351804 Mus muscu	
	12	270.6	13.4	1378	10	F351799S11		AF351809 Mus muscu	
	13	264.8	13.1	237445	2	AC120701		AC120701 Rattus no	
c	14	264.8	13.1	312858	2	AC112747		AC112747 Rattus no	
c	15	264.2	13.1	40929	10	AY145899		AY145899 Rattus no	
	16	245.2	12.1	1470	10	F351799S04		AF351802 Mus muscu	
	17	225.8	11.2	207760	2	AC146286		AC146286 Callicebu	
	18	225.6	11.2	660	9	F351812S06		AF351817 Homo sapi	
	19	225.6	11.2	127066	9	AC084265		AC084265 Homo sapi	
	20	225.6	11.2	139342	9	AC108476		AC108476 Homo sapi	
c	21	222.4	11.0	202533	2	AC146464		AC146464 Saimiri s	
	22	219.4	10.9	178016	2	AC146787		AC146787 Aotus nan	
	23	219.4	10.9	185045	2	AC146466		AC146466 Callithri	
c	24	216.8	10.7	159346	2	AC145533		AC145533 Lemur cat	
c	25	212	10.5	68166	2	AC084712		AC084712 Homo sapi	
	26	206.2	10.2	1292	9	F351812S11		AF351822 Homo sapi	
	27	205.2	10.2	642	10	F351799S09		AF351807 Mus muscu	
	28	205.2	10.2	182261	2	AC087053		AC087053 Homo sapi	
	29	199.2	9.9	1920	6	AX456519		AX456519 Sequence	
	30	199.2	9.9	2340	6	AX320883		AX320883 Sequence	
	31	199.2	9.9	2340	6	AX685733		AX685733 Sequence	
	32	199.2	9.9	2340	9	AF320293		AF320293 Homo sapi	
	33	199.2	9.9	2516	6	AX456520		AX456520 Sequence	

34	199.2	9.9	2740	9	AF312715	AF312715 Homo sapi
35	195	9.7	2035	6	AX456526	AX456526 Sequence
36	195	9.7	2470	10	AF312714	AF312714 Rattus no
37	189.4	9.4	159346	2	AC145533	AC145533 Lemur cat
38	188.2	9.3	2351	10	AY195872	AY195872 Mus muscu
39	188.2	9.3	2351	10	AY195873	AY195873 Mus muscu
40	186.6	9.2	1915	6	AX456523	AX456523 Sequence
41	186.6	9.2	1959	6	AX685729	AX685729 Sequence
42	186.6	9.2	2258	6	AX320881	AX320881 Sequence
43	186.6	9.2	2354	6	AX456524	AX456524 Sequence
44	186.6	9.2	2354	10	AF312713	AF312713 Mus muscu
c 45	185	9.2	64889	2	AC120532	AC120532 Oryza sat

ALIGNMENTS

RESULT 1

AX685731

LOCUS AX685731 2019 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 3 from Patent WO02081691.

ACCESSION AX685731

VERSION AX685731.1 GI:29371740

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 3 17-OCT-2002;
Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)

FEATURES

source

Location/Qualifiers

1. .2019

/organism="Mus musculus"

/mol_type="unassigned DNA"

/db_xref="taxon:10090"

CDS

1. .2019

/note="unnamed protein product; mouse ABCG8 (mABCG8)"

/codon_start=1

/protein_id="CAD86571.1"

/db_xref="GI:29371741"

/db_xref="REMTREMBL:CAD86571"

/translation="MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQS
NTLEVRDLTYQVDIASQVPWFEOQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLA
IIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKCAVHRQHDQLLPN
LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
RRVSIGVQLLWNPGLILIDEPSTGLDSFTAHNLVTTLSRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGAAQQM VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKER
EVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTTLTQDTCGTAVELP
GMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAAL
LFMIGALIPFNVILDVVS KCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVI
IYAMP IYWL TNLRPVP ELLHFLLVVLFCCRTMALAASAMLP TFHMS SFFCNALY
NSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSIL
GDTMISAMDLNSHPLYAIYLIVIGISYGFLLYYLSLKLKIQKSIQDW"

ORIGIN

Query Match 100.0%; Score 2019; DB 6; Length 2019;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60

Qy     61 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120

Qy    121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180

Qy    181 CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240

Qy    241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300

Qy    301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360

Qy    361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420

Qy    421 CCTCAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCCAAC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTCAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCCAAC 480

Qy    481 CTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC 540

Qy    541 CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGTC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGTC 600

Qy    601 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA 660

Qy    661 GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 720

Qy    721 TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTCGCCTGGCCAAG 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTCGCCTGGCCAAG 780
```

Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC	960
Db	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC	960
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Db	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Qy	1021	AAGGCACAGTCTCTTGCAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1021	AAGGCACAGTCTCTTGCAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Qy	1081	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1081	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Qy	1141	CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTCCACC	1200
Db	1141	CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTCCACC	1200
Qy	1201	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCATGGG	1260
Db	1201	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCATGGG	1260
Qy	1261	TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Db	1261	TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Qy	1321	CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Db	1321	CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Qy	1381	AATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1440
Db	1381	AATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1440
Qy	1441	CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Db	1441	CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Qy	1501	CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Db	1501	CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Qy	1561	CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620
Db	1561	CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620

Qy 1621 TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC 1680
 |||
 Db 1621 TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC 1680
 Qy 1681 TTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC 1740
 |||
 Db 1681 TTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC 1740
 Qy 1741 AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTCTCG 1800
 |||
 Db 1741 AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTCTCG 1800
 Qy 1801 GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC 1860
 |||
 Db 1801 GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC 1860
 Qy 1861 TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT 1920
 |||
 Db 1861 TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT 1920
 Qy 1921 GCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTATCC 1980
 |||
 Db 1921 GCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTATCC 1980
 Qy 1981 TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
 |||
 Db 1981 TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019

RESULT 2

AY196216

LOCUS AY196216 2284 bp mRNA linear ROD 01-JUN-2003

DEFINITION Mus musculus strain PERA/Ei ATP-binding cassette sub-family G member 8 (Abcg8) mRNA, complete cds.

ACCESSION AY196216

VERSION AY196216.1 GI:31322261

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2284)

AUTHORS Wittenburg,H., Lyons,M.A., Li,R., Churchill,G.A., Carey,M.C. and Paigen,B.

TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred Mice

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2284)

AUTHORS Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.

TITLE Direct Submission

JOURNAL Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

FEATURES Location/Qualifiers

source 1..2284

/organism="Mus musculus"

/mol_type="mRNA"

```

/strain="PERA/Ei"
/db_xref="taxon:10090"
/chromosome="17"
/map="55 cM"
/sex="male"
/tissue_type="liver"
gene 1. .2284
/ gene="Abcg8"
CDS 102. .2120
/ gene="Abcg8"
/ note="ATP-dependent canalicular cholesterol transporter;
white subfamily"
/ codon_start=1
/ product="ATP-binding cassette sub-family G member 8"
/ protein_id="AAO45096.1"
/ db_xref="GI:31322262"
/ translation="MAEKTKEETQLWNGTVLQDASGLQDSLFSSSDNSLYFTYSGQS
NTLEVRDLTYQVDIASQVPWFQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQGMLA
IIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPN
LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKER
EVATVEKAQSLAALFLEKVGQGFDDFLWKAELKELNTSTHTVSLTLTQDTCGTAVELP
GMIEQFSTLIRRIISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAAL
LFMIGALIPFNVILDVSKCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVI
IYAMPIYWLTNLRPVPELFLHFLLVVLFVFCRTMALAASAMLPFHMSSFFCNALY
NSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSIL
GDTMISAMDLNSHPLYAIYLVIGISYGFLLFLYYLSLKLKQKSIQDW"

```

ORIGIN

```

Query Match          100.0%;  Score 2019;  DB 10;  Length 2284;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 2019;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 ATGGCTGAGAAAACCAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      102 ATGGCTGAGAAAACCAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 161

Qy      61 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      162 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 221

Qy      121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      222 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 281

Qy      181 CAGGTGCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      282 CAGGTGCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 341

Qy      241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      342 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 401

Qy      301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      402 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 461

```

Qy 361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420
 |||
 Db 462 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 521

Qy 421 CCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCCAAC 480
 |||
 Db 522 CCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCCAAC 581

Qy 481 CTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC 540
 |||
 Db 582 CTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC 641

Qy 541 CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC 600
 |||
 Db 642 CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC 701

Qy 601 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA 660
 |||
 Db 702 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA 761

Qy 661 GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 720
 |||
 Db 762 GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 821

Qy 721 TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCGCTGGCCAAG 780
 |||
 Db 822 TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCGCTGGCCAAG 881

Qy 781 GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT 840
 |||
 Db 882 GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT 941

Qy 841 GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG 900
 |||
 Db 942 GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG 1001

Qy 901 GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC 960
 |||
 Db 1002 GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC 1061

Qy 961 TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG 1020
 |||
 Db 1062 TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG 1121

Qy 1021 AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG 1080
 |||
 Db 1122 AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG 1181

Qy 1081 TGGAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCTCACA 1140
 |||
 Db 1182 TGGAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCTCACA 1241

Qy 1141 CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCC GGATGATAGAGCAGTTTCCACC 1200
 |||
 Db 1242 CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCC GGATGATAGAGCAGTTTCCACC 1301

Qy	1201	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCATGGG	1260
Db	1302	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCATGGG	1361
Qy	1261	TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Db	1362	TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1421
Qy	1321	CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Db	1422	CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1481
Qy	1381	AATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1440
Db	1482	AATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1541
Qy	1441	CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Db	1542	CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1601
Qy	1501	CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Db	1602	CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1661
Qy	1561	CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620
Db	1662	CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1721
Qy	1621	TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC	1680
Db	1722	TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC	1781
Qy	1681	TTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC	1740
Db	1782	TTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC	1841
Qy	1741	AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTCTCG	1800
Db	1842	AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTCTCG	1901
Qy	1801	GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC	1860
Db	1902	GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC	1961
Qy	1861	TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCCACTCTAT	1920
Db	1962	TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCCACTCTAT	2021
Qy	1921	GCGATCTACCTCATTTGTCATCGGCATCAGCTACGGCTTCCTGTTCCCTGTACTATCTATCC	1980
Db	2022	GCGATCTACCTCATTTGTCATCGGCATCAGCTACGGCTTCCTGTTCCCTGTACTATCTATCC	2081
Qy	1981	TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019
Db	2082	TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2120

RESULT 3
 AF324495
 LOCUS AF324495 3674 bp mRNA linear ROD 07-AUG-2001
 DEFINITION Mus musculus sterolin-2 (Abcg8) mRNA, complete cds.
 ACCESSION AF324495
 VERSION AF324495.1 GI:15088541
 KEYWORDS .
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 3674)
 AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H.,
 Ose,L., Stalenhoef,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E.,
 Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and
 Patel,S.B.
 TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterolin-1 and
 sterolin-2, encoded by ABCG5 and ABCG8, respectively
 JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
 MEDLINE 21344600
 PUBMED 11452359
 REFERENCE 2 (bases 1 to 3674)
 AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 Street, STB541, Charleston, SC 29403, USA
 FEATURES Location/Qualifiers
 source 1. .3674
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /tissue_type="liver"
 gene 1. .3674
 /gene="Abcg8"
 CDS 102. .2123
 /gene="Abcg8"
 /note="ABCG8"
 /codon_start=1
 /product="sterolin-2"
 /protein_id="AAK84079.1"
 /db_xref="GI:15088542"
 /translation="MAEKTKEETQLWNGTVLQDASQGLQDSLFSSSESDNSLYFTYSGQ
 SNTLEVRDLTYQVDIASQVPWFEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQML
 AIIGSSGCGRASLLDVITGRGHGGMKMSGQIWINQGPSTPQLVRKCVAVHRQHDQLLP
 NLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGE
 RRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSD
 IFRFLDLVLLMTSGTPPIYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKE
 REVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTLTQDPTDCGTAVEL
 PGMIEQFSTLIRRQISNDFRDLPDLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAA
 LLFMIGALIPFNVILDVVSCKHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYV
 IYAMPYIYWLTLNLRPPELFLHLLVWLVVFCRTMALAASAMLPFHMSSFECNAL
 YNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSI
 LGDTMISAMDLNSHPLYAIYLIVIGISYGFLLFLYYLSLKLKQKSIQDW"

ORIGIN

Query Match 99.4%; Score 2006; DB 10; Length 3674;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2019; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```
Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 161

Qy      61 TC---GGGCCTCCAGGACAGCTTGTTCCTCTCGGAAAGTGACAACAGTCTGTACTTCACC 117
          || ||||||||||||||||||||||||||||||||||||||||||||
Db     162 TCGCAGGGCCTCCAGGACAGCTTGTTCCTCTCGGAAAGTGACAACAGTCTGTACTTCACC 221

Qy     118 TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC 177
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     222 TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC 281

Qy     178 TCTCAGGTGCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATACCTGGAGGTCTCATAGC 237
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     282 TCTCAGGTGCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATACCTGGAGGTCTCATAGC 341

Qy     238 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 297
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     342 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 401

Qy     298 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 357
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     402 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 461

Qy     358 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 417
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     462 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 521

Qy     418 ACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCC 477
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     522 ACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCC 581

Qy     478 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTC 537
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     582 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTC 641

Qy     538 TCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAG 597
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     642 TCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAG 701

Qy     598 TGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGA 657
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     702 TGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGA 761

Qy     658 CGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCC 717
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     762 CGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCC 821

Qy     718 ACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTCGCCTGGCC 777
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     822 ACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTCGCCTGGCC 881
```

Qy 778 AAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTA 837
 |||
 Db 882 AAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTA 941

Qy 838 TTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAA 897
 |||
 Db 942 TTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAA 1001

Qy 898 ATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGAC 957
 |||
 Db 1002 ATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGAC 1061

Qy 958 TTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTG 1017
 |||
 Db 1062 TTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTG 1121

Qy 1018 GAGAAGGCACAGTCTCTTGACAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTT 1077
 |||
 Db 1122 GAGAAGGCACAGTCTCTTGACAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTT 1181

Qy 1078 CTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTC 1137
 |||
 Db 1182 CTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTC 1241

Qy 1138 ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCC 1197
 |||
 Db 1242 ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCC 1301

Qy 1198 ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT 1257
 |||
 Db 1302 ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT 1361

Qy 1258 GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC 1317
 |||
 Db 1362 GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC 1421

Qy 1318 AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT 1377
 |||
 Db 1422 AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT 1481

Qy 1378 TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT 1437
 |||
 Db 1482 TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT 1541

Qy 1438 GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA 1497
 |||
 Db 1542 GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA 1601

Qy 1498 TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC 1557
 |||
 Db 1602 TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC 1661

Qy 1558 CTGCGGCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC 1617
 |||
 Db 1662 CTGCGGCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC 1721

Qy 1618 TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCCTCC 1677

```

|||||
Db      1722 TGTCGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTCCACATGTCCTCC 1781
Qy      1678 TTCTTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAACTTG 1737
|||||
Db      1782 TTCTTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAACTTG 1841
Qy      1738 GACAACTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTC 1797
|||||
Db      1842 GACAACTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTC 1901
Qy      1798 TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC 1857
|||||
Db      1902 TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC 1961
Qy      1858 ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC 1917
|||||
Db      1962 ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC 2021
Qy      1918 TATGCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA 1977
|||||
Db      2022 TATGCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA 2081
Qy      1978 TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
|||||
Db      2082 TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2123

```

RESULT 4

AY196215

LOCUS AY196215 2285 bp mRNA linear ROD 01-JUN-2003

DEFINITION Mus musculus strain I/LnJ ATP-binding cassette sub-family G member 8 (Abcg8) mRNA, complete cds.

ACCESSION AY196215

VERSION AY196215.1 GI:31322259

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2285)

AUTHORS Wittenburg,H., Lyons,M.A., Li,R., Churchill,G.A., Carey,M.C. and Paigen,B.

TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred Mice

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2285)

AUTHORS Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.

TITLE Direct Submission

JOURNAL Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

FEATURES Location/Qualifiers

```

source      1. .2285
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="I/LnJ"

```

```

/db_xref="taxon:10090"
/chromosome="17"
/map="55 cM"
/sex="male"
/tissue_type="liver"
gene 1. .2285
/gene="Abcg8"
CDS 102. .2120
/gene="Abcg8"
/note="ATP-dependent canalicular cholesterol transporter;
white subfamily"
/codon_start=1
/product="ATP-binding cassette sub-family G member 8"
/protein_id="AAO45095.1"
/db_xref="GI:31322260"
/translation="MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQS
NTLEVRDLTYQVDIASQVPWFQQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLA
IIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPN
LTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLI SLHQPRSDI
FRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKER
EVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTLTQDTCGTAAELP
GMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAAL
LFMIGALIPFNVIDVVSCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVI
IYAMPIYWLTLNLRVPPELFLHLLLVLVVFCRTMALAASAMLPTFHMSSFFCNALY
NSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSIL
GDTMISAMDNLNSHPLYAIYLIVIGISYGFLEFLYYLSLKLKQKSIQDW"

```

ORIGIN

```

Query Match          99.0%;  Score 1999.8;  DB 10;  Length 2285;
Best Local Similarity 99.4%;  Pred. No. 0;
Matches 2007;  Conservative 0;  Mismatches 12;  Indels 0;  Gaps 0;

Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      102 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 161

Qy      61 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      162 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 221

Qy      121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
        || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      222 AGCGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 281

Qy      181 CAGGTGCCTTGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      282 CAGGTGCCTTGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 341

Qy      241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      342 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 401

Qy      301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      402 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 461

```

Qy	361	AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG	420
Db	462	AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAACGGGCAACCCAGTACG	521
Qy	421	CCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAAC	480
Db	522	CCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAAC	581
Qy	481	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC	540
Db	582	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC	641
Qy	541	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	642	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	701
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	660
Db	702	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	761
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Db	762	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	821
Qy	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCCCTGGCCAAG	780
Db	822	TCTGGCCTCGACAGCTTCACAGCCCACAACCTGGTGACAACCTTGTCGCCCTGGCCAAG	881
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	882	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	941
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	942	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	1001
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC	960
Db	1002	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC	1061
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Db	1062	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1121
Qy	1021	AAGGCACAGTCTCTTGAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1122	AAGGCACAGTCTCTTGAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1181
Qy	1081	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1182	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1241
Qy	1141	CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTCCACC	1200
Db	1242	CAGGACACTGACTGTGGGACTGCTGCTGAGCTGCCCCGGGATGATAGAGCAGTTTCCACC	1301
Qy	1201	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCATGGG	1260

Db	1302	 CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTTCATGGG	1361
Qy	1261	TCGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Db	1362	 TCGGAAGCCTGCCTGATGTCCCTCATCATCGGCTTCCTTTACTACGGCCATGGGGCCAAG	1421
Qy	1321	CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Db	1422	 CAGCTCTCCTTCATGGACACGGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1481
Qy	1381	AATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1440
Db	1482	 AATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1541
Qy	1441	CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Db	1542	 CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1601
Qy	1501	CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Db	1602	 CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1661
Qy	1561	CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620
Db	1662	 CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTACTGCTTGTGTGGTTGGTGGTCTTCTGC	1721
Qy	1621	TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC	1680
Db	1722	 TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC	1781
Qy	1681	TTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC	1740
Db	1782	 TTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACCGCCGGCTTCATGATAAACTTGGAC	1841
Qy	1741	AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTCTCG	1800
Db	1842	 AACCTGTGGATAGTGCCTGCATGGATATCCAAGCTGTCGTTCTCCGGTGGTGCTTCTCG	1901
Qy	1801	GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC	1860
Db	1902	 GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC	1961
Qy	1861	TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT	1920
Db	1962	 TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT	2021
Qy	1921	GCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTATCC	1980
Db	2022	 GCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTATCC	2081
Qy	1981	TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019
Db	2082	 TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2120

RESULT 5

AF351785
LOCUS AF351785 4829 bp mRNA linear ROD 26-AUG-2002
DEFINITION Rattus norvegicus sterolin-2 (Abcg8) mRNA, complete cds.
ACCESSION AF351785
VERSION AF351785.2 GI:22477145
KEYWORDS .
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4829)
AUTHORS Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidaka,H., Kojima,H.,
Ose,L., Stalenhoeft,A.F., Mietinnen,T., Bjorkhem,I., Bruckert,E.,
Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Srivastava,A. and
Patel,S.B.
TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterolin-1 and
sterolin-2, encoded by ABCG5 and ABCG8, respectively
JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
MEDLINE 21344600
PUBMED 11452359
REFERENCE 2 (bases 1 to 4829)
AUTHORS Lu,K., Yu,H., Lee,M. and Patel,S.B.
TITLE Molecular cloning, genomic structure, and characterization of novel
mouse head-to-head tandem ABC transporters
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 4829)
AUTHORS Lu,K., Lee,M. and Patel,S.B.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29407, USA
REFERENCE 4 (bases 1 to 4829)
AUTHORS Lu,K., Yu,H., Lee,M. and Patel,S.B.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-2002) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403, USA
REMARK Sequence update by submitter
COMMENT On Aug 26, 2002 this sequence version replaced gi:15148516.
FEATURES
source Location/Qualifiers
1. .4829
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
gene 1. .4829
/gene="Abcg8"
CDS 111. .2129
/gene="Abcg8"
/codon_start=1
/product="sterolin-2"
/protein_id="AAK84831.2"
/db_xref="GI:22477146"
/translation="MAEKTKEETQLWNGTVLQDASSLQDSVFSSESNDNSLYFTYSGQS
NTLEVRDLTYQVDMASQVPWFEEQLAQFKLPWRSRGSQDSWDLGIRNLSFKVRSQGMLA

IIGSAGCGRATLLDVITGRDHGGKMKSGQIWINGQPSTPQLIQKCVAHVRQDQDLLPN
 LTVRETLTFIAQMRLPKTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
 RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVRTLSRLAKGNRLVLI SLHQPRSDI
 FRLFDLVLLMTSGTPIYLGVAQHMVQYFTSIGYPCPRYSNPADFYVDLTSIDRRSKEQ
 EVATMEKARLLAALFLEKVQGFDDFLWKAEAKSLDTGTYAVSQTILTQDTNCGTAAELP
 GMIQQFTTLIRRQISNDFRDLPTLFHGAELCLMSLIIGFLYYGHADKPLSFMDMAAL
 LFMIGALIPFNVILDVVSCHSERSLLYYELEDGLYTAGPYFFAKVLGELPEHCAYVI
 IYGMPIYWLTNLRPGPELFLHFMLLWLTVFCCRTMALAASAMLPTFHMSSFCCNALY
 NSFYLTAGFMINLNNLWIVPAWISKMSFLRWCFSGLMQIQFNHGYTTQIGNLTFSVF
 GDAMVTAMDNLNSHPLYAIYLIVIGISCGFLSLYYLSLKFQKSIQDW"

ORIGIN

Query Match 85.6%; Score 1727.8; DB 10; Length 4829;
 Best Local Similarity 91.0%; Pred. No. 0;
 Matches 1837; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

Qy	1	ATGGCTGAGAAAACCAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT	60
Db	111	ATGGCTGAGAAGACCAAGAGAGACCCAGCTGTGGAACGGGACTGTACTCCAGGATGCT	170
Qy	61	TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC	120
Db	171	TCAAGCCTCCAGGACAGCGTGTTCTCCTCTGAAAGTGACAACAGCCTCTACTTCACCTAC	230
Qy	121	AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT	180
Db	231	AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATGGCCTCT	290
Qy	181	CAGGTGCCTTGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC	240
Db	291	CAGGTGCCTTGTTTGAGCAGCTGGCTCAGTTCAAGTTACCGTGGAGGTCTCGCGGCAGC	350
Qy	241	CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG	300
Db	351	CAGACTCCTGGGATCTGGGCATCCGAAATCTGAGCTTCAAAGTGAGGAGTGGACAGATG	410
Qy	301	CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC	360
Db	411	CTGGCTATCATAGGGAGCGCAGGCTGCGGGAGAGCCACATTACTCGACGTTATCACAGGC	470
Qy	361	AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG	420
Db	471	AGAGACCATGGTGGCAAGATGAAATCAGGACAAATCTGGATAAACGGGCAACCCAGCAGC	530
Qy	421	CCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAAC	480
Db	531	CCTCAGCTGATACAGAAGTGTGTGGCACATGTGCGCCAGCAAGACCAGTGCTCCCAAT	590
Qy	481	CTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC	540
Db	591	CTGACTGTCAGAGAGACCCTGACTTTTCATCGCCCAGATGCGCCTGCCAAGACCTTCTCT	650
Qy	541	CAGGCCACGCTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	651	CAGGCCACGCGAGACAAACGGGTGGAAGACGTGATTGCGGAGCTGCGGCTGCGGCAGTGC	710
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGTGAGCGCCGACGA	660

Db	711	GCCAACACCCGCGTGGGCAACACATACGTACGCGGGGTGTCCGGGGGCGAGCGCCGAAGA	770
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Db	771	GTGAGCATCGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATCCTGGATGAACCCACT	830
Qy	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAG	780
Db	831	TCCGGCCTCGACAGCTTCACCGCTCACAACCTGGTGAGAACTTTGTCCCGCCTGGCCAAA	890
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	891	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	950
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	951	GACCTGGTCCTTCTGATGACGTCTGGCACCCTATCTACCTGGGGGTGGCACAGCACATG	1010
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCTCGCTATAGCAACCCTGCGGACTTC	960
Db	1011	GTGCAGTACTTTACATCAATTGGCTACCCTTGTCTCTCGCTACAGCAACCCTGCTGACTTC	1070
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Db	1071	TACGTGGACTTGACGAGCATTGACAGGCGCAGCAAAGAACAGGAGGTGGCCACCATGGAG	1130
Qy	1021	AAGGCACAGTCTCTTGAGCCCTGTTCTTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1131	AAGGCTCGATTACTTGACGCCTTGTTCTTAGAAAAAGTGCAAGGCTTTGACGACTTTCTG	1190
Qy	1081	TGGAAAGCTGAGGCAAAGGAAGTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1191	TGGAAAGCTGAGGCAAAGAGTCTCGACACAGGCACCTATGCAGTCAGCCAGACCCTCACA	1250
Qy	1141	CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGATGATAGAGCAGTTTCCACC	1200
Db	1251	CAGGACACCAACTGTGGAAGTGTGCTGAGCTGCCCGGATGATACAGCAGTTTACCACC	1310
Qy	1201	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTTCATGGG	1260
Db	1311	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACTTGCCACCCCTGTTTCATCCATGGA	1370
Qy	1261	TCGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Db	1371	GCAGAAGCCTGCCTGATGTCTCTCATCATTTGGCTTCCTTTACTACGGCCACGCAGACAAG	1430
Qy	1321	CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTTCCTTC	1380
Db	1431	CCGCTCTCCTTCATGGACATGGCAGCCCTCCTGTTTCATGATAGGAGCACTCATTTCCTTT	1490
Qy	1381	AATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1440
Db	1491	AATGTCATTCTGGATGTCGTCTCCAAATGTCACTCGGAGCGGTGCTGCTGTACTATGAA	1550
Qy	1441	CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500

Db 1551 CTGGAGGACGGACTGTACACTGCTGGTCCTTATTTCTTTGCCAAGGTCCTCGGTGAGCTG 1610

Qy 1501 CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG 1560
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1611 CCAGAGCACTGTGCCTATGTCATCATCTATGGGATGCCCATCTACTGGCTGACCAACCTG 1670

Qy 1561 CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC 1620
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1671 CGGCCAGGGCCTGAGCTCTTCCTCCTGCACTTCATGCTTCTGTGGCTGGTGGTGTCTGC 1730

Qy 1621 TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC 1680
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1731 TGCAGGACCATGGCCCTGGCCGCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC 1790

Qy 1681 TTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC 1740
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1791 TGCTGCAACGCTCTCTACAACCTCTTCTACCTTACGGCTGGCTTCATGATAAACTTGAAC 1850

Qy 1741 AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCCTTCTCG 1800
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1851 AACCTGTGGATAGTACCTGCATGGATTTCGAAGATGTCGTTCTCCGGTGGTGCCTTCTCA 1910

Qy 1801 GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC 1860
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1911 GGGCTGATGCAGATTCAATTTAATGGACACATTTACACCACGCAGATCGGCAACCTCACC 1970

Qy 1861 TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT 1920
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1971 TTCTCCGTCCCCGGAGACGCGATGGTCACTGCCATGGACCTGAACTCACATCCTCTTTAT 2030

Qy 1921 GCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCCTGTTCCCTGTACTATCTATCC 1980
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2031 GCGATCTACCTCATCGTCATTGGCATCAGCTGTGGCTTCCTGTCCCTGTATTATCTGTCC 2090

Qy 1981 TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2091 TTGAAGTTCATCAAACAGAAGTCAATTCAAGATTGGTGA 2129

RESULT 6

AX685735

LOCUS AX685735 2669 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 7 from Patent WO02081691.

ACCESSION AX685735

VERSION AX685735.1 GI:29371744

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 7 17-OCT-2002;

Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
 (US)

FEATURES

Location/Qualifiers

```

source      1.2669
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

CDS         100..2121
            /note="unnamed protein product; human ABCG8 (hABCG8)"
            /codon_start=1
            /protein_id="CAD86573.1"
            /db_xref="GI:29371745"
            /db_xref="REMTREMBL:CAD86573"
            /translation="MAGKAAEERGLPKGATPQDTSGLQDRLFSSSEDNSLYFTYSGQP
NTLEVRDLNYQVDLASQVPWFQEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQGMLA
IIGSSGCGRASLLDVITGRGHGGKIKSGQIWIWNGQPSSPQLVRKCAHVQRHQNQLLPN
LTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER
RRVSIQVQLLWNPGLILILDEPTSGLDSTAHNLVKTLSRLAKGNRLVLI SLHQPRSDI
FRLFDLVLLMTSGTPPIYLGAQAHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQ
ELATREKAQSLAALFLEKVRDLDDFLWKAETKDLDDETCVESSVTPLDTNCLPSPTKM
PGAVQQQFTTLIRROI SNDFRDLPTLLIHGAEACLSMTIGFLYFGHSGSIQLSFMDTAA
LLFMIGALIPFNVLVDVISKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYI
IIYGMPTYWLANLRPGLQPPELLHFLLVLVVFCRIMALAAALLPTEFHMAFFSNAL
YNSFYLAGGFMINLSSLWTPPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAV
SGDKILSAMELDSYPLYAIYLVIVIGLSSGGFMVLYYVSLRFIKQKPSQDW"

```

Query Match 70.8%; Score 1430; DB 6; Length 2669;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 1659; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

Db	520	 CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACAGCTGCTCCCCAAC	579
Qy	481	CTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCAGATGCGCCTGCCAGGACCTTCTCC	540
Db	580	TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCAGATGCGGCTGCCAGAACCTTCTCC	639
Qy	541	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	640	CAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGC	699
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGTGAGCGCCGACGA	660
Db	700	GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGTGTGTCGGGGGTGAGCGCAGGAGA	759
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTTCTGGATGAACCCACT	720
Db	760	GTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC	819
Qy	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCTGGCCAAAG	780
Db	820	TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAA	879
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	880	GGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTT	939
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	940	GATCTGGTCCTCCTGATGACGTCTGGCACCCCCATCTACTTAGGGGCGGCCAGCACATG	999
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCTGCGGACTTC	960
Db	1000	GTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTC	1059
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCCTGGAG	1020
Db	1060	TATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAG	1119
Qy	1021	AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1120	AAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCCTGACTTAGATGACTTTCTA	1179
Qy	1081	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1180	TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCA	1239
Qy	1141	CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTTC	1197
Db	1240	CTAGACACCAACTGCCTCCCGAGTCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG	1299
Qy	1198	ACCCTGATCCGTCGTGAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT	1257
Db	1300	ACGCTGATCCGTCGTGAGATTTCCAACGACTTCCGAGACCTGCCACCCTCCTCATCCAT	1359
Qy	1258	GGGTGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1317

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2022)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J.,
Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Accumulation of Dietary Cholesterol in Sitosterolemia Caused by
Mutations in Adjacent ABC Transporters

JOURNAL Science (2001) In press

REFERENCE 2 (bases 1 to 2022)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J.,
Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Direct Submission

JOURNAL Submitted (09-NOV-2000) Molecular Genetics, University of Texas,
Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd.,
Dallas, TX 75390-9046, USA

FEATURES Location/Qualifiers

source 1. .2022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

gene 1. .2022
/gene="ABCG8"

CDS 1. .2022
/gene="ABCG8"
/note="ATP-binding cassette, subfamily G, member 8"
/codon_start=1
/product="ABCG8"
/protein_id="AAG40004.1"
/db_xref="GI:11692802"
/translation="MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQP
NTLEVRDLNYQVDLASQVPWFELAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLA
IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSPQLVRKCVAVHRQHNQLLPN
LTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGAAQHMVQYFATAIGYPCPRYSNPADFYVDLTSIDRRSREQ
ELATREKAQSLAALFLEKVRDLDDFLWKAETKDLEDETCVESSVTPLDNTCLPSPPTKM
PGAVQQFTTLIRQISNDFRDLPTLLIHGAELMSMTIGFLYFGHGSIQLSFMDTAA
LLEFMIGALIPFNVILDVISKYSERAMLYELEDGLYTTGPYFFAKILGELPEHCAYI
IIYGMPTYWLANLRPGLQPFLLHFLLVLVVFCCRIMALAAAALLPTFHMASFSSNAL
YNSFYLAGGEMINLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPPLGNLTIAV
SGDKILSVMELDSYPLYAIYLIVIGLSGGFMVLYYVSLRFIKQKPSQDW"

ORIGIN

Query Match 70.7%; Score 1428.4; DB 9; Length 2022;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 1658; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

Qy	1	ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT	60
Db	1	ATGGCCGGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGCCACTCCCCAGGATACC	60
Qy	61	TCGGGCCTCCAGGACAGCTTGTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC	120
Db	61	TCGGGCCTCCAGGATAGATTGTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTAC	120

Qy	121	AGTGGTCAGTCCAACACTCTGGAGGTGAGAGATCTCACCTACCAGGTGGACATCGCCTCT	180
Db	121	AGTGGCCAGCCCAACACCCTGGAGGTGAGAGACCTCAACTACCAGGTGGACCTGGCCTCT	180
Qy	181	CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC	240
Db	181	CAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGC	240
Qy	241	CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG	300
Db	241	CAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATG	300
Qy	301	CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC	360
Db	301	CTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGC	360
Qy	361	AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG	420
Db	361	CGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCG	420
Qy	421	CCTCAGCTGGTGAGGAAGTGCCTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCAAC	480
Db	421	CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACAGCTGCTCCCCAAC	480
Qy	481	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC	540
Db	481	TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCC	540
Qy	541	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	541	CAGGCCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGC	600
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	660
Db	601	GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAGCGCAGGAGA	660
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Db	661	GTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC	720
Qy	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCTGGCCAAG	780
Db	721	TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAA	780
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	781	GGCAACCGGCTGGTGCTCATCTCCCTCCACAGCCTCGCTCTGACATCTTCAGGCTGTTT	840
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	841	GATCTGGTCCTCCTGATGACGTCTGGCACCCCACTCTACTTAGGGGCGGCCAGCACATG	900
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCTGCGGACTTC	960
Db	901	GTCCAGTATTTTCACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCTGCTGACTTC	960

Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Db	961	TATGTGGACCTGACCAGCATTTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCAGGGAG	1020
Qy	1021	AAGGCACAGTCTCTTGCAGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTTCTG	1080
Db	1021	AAGGCTCAGTCACTCGCAGCCCTGTTCTAGAAAAAGTGCCTGACTTAGATGACTTTCTA	1080
Qy	1081	TGGAAAGCTGAGGCAAAGGAACCTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1081	TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCA	1140
Qy	1141	CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTTC	1197
Db	1141	CTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG	1200
Qy	1198	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT	1257
Db	1201	ACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGCCACCTCCTCATCCAT	1260
Qy	1258	GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTTGGCTTCCTTTACTACGGCCATGGGGCC	1317
Db	1261	GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGC	1320
Qy	1318	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1377
Db	1321	ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCT	1380
Qy	1378	TTCAATGTCTCCTGGATGTGCTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT	1437
Db	1381	TTCAACGTCTCTGGATGTCTCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT	1440
Qy	1438	GAGCTGGAAGACGGGCTGTACACTGCTGGTCTTATTTCTTTGCCAAGATCCTAGGAGAA	1497
Db	1441	GAAGTGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG	1500
Qy	1498	TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1557
Db	1501	CTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAAC	1560
Qy	1558	CTGCGGCCCCGTGCCTGAGCTCTTCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC	1617
Db	1561	CTGAGGCCAGGCCTCCAGCCCTTCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC	1620
Qy	1618	TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCTCTCC	1677
Db	1621	TGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCC	1680
Qy	1678	TTCTTCTGCAATGCCCTCTACAACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG	1737
Db	1681	TTCTTCAGCAATGCCCTCTACAACTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG	1740
Qy	1738	GACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTGCTTCCTCCGGTGGTGGTCTTC	1797
Db	1741	AGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCCTGCGGTGGTGGTCTTC	1800
Qy	1798	TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC	1857

Db	1801	GAAGGGCTGATGAAGATTCTAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTC	1860
Qy	1858	ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCCACTC	1917
Db	1861	ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGTCATGGAGCTGGACTCGTACCCTCTC	1920
Qy	1918	TATGCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCCCTGTACTATCTA	1977
Db	1921	TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTG	1980
Qy	1978	TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019
Db	1981	TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA	2022

AF324494

LOCUS AF324494 2679 bp mRNA linear PRI 07-AUG-2001

DEFINITION Homo sapiens sterolin-2 (ABCG8) mRNA, complete cds.

ACCESSION AF324494

VERSION AF324494.1 GI:15088539

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2679)

AUTHORS Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H., Ose, L., Stalenhoef, A.F., Mietinnen, T., Bjorkhem, I., Bruckert, E., Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Srivastava, A. and Patel, S.B.

TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)

MEDLINE 21344600

PUBMED 11452359

REFERENCE 2 (bases 1 to 2679)

AUTHORS Lu, K., Lee, M.-H. and Patel, S.B.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty Street, STB541, Charleston, SC 29403, USA

FEATURES	Location/Qualifiers
----------	---------------------

source

1. .2679

```
/organism="Homo sapiens"
```

```
/mol type="mRNA"
```

```
/db xref="taxon:9606"
```

```
/chromosome="2"
```

```
/map="2p21; between D2S2294 and D2S2298"
```

```
/tissue type="liver"
```

gene

$$1. \quad .267\overline{9}$$

```
/gene="ABCG8"
```

CDS

91. .2112

```
/gene="ABCG8"
```

```
/codon start=1
```

```

/product="sterolin-2"
/protein_id="AAK84078.1"
/db_xref="GI:15088540"
/translation="MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQP
NTLEVRDLNCQVDLASQVPWFEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLA
IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVRKCVAVHRQHNQLLPN
LTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER
RRVSIQVQLLWNP GILILDEPTSGLDSFTAHLNVLKTLRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQ
ELATREKAQSLAALFLEKVRDLDDFLWKAETKDLEDTCVESSVTPLDTNCLPSPTKM
PGAVQQFTTLIRRQISNDFRDLPTLLIHGAELMSMTIGFLYFGHGSIQLSFMDTAA
LLFMIGALIPFNVILDVISKYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYI
IIYGMPTYWLANLRPGLQPFLHFLLVLVVFCCRIMALAAAALLPTFHMASFFSNAL
YNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMP LGNLTIAV
SGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYVSLRFIKQKPSQDW"

```

ORIGIN

```

Query Match          70.7%;  Score 1428.4;  DB 9;  Length 2679;
Best Local Similarity 82.0%;  Pred. No. 0;
Matches 1658;  Conservative 0;  Mismatches 361;  Indels 3;  Gaps 1;

```

```

Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||| | ||| |  ||| ||||  |||| ||| ||| |  | ||||| |
Db      91 ATGGCCGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACC 150

Qy      61 TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120
        ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     151 TCGGGCCTCCAGGATAGATTGTTCTCCTCTGAAAGTGACAACAGCCTGTACTTCACCTAC 210

Qy     121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
        ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     211 AGTGGCCAGCCCAACACCCTGGAGGTCAGAGACCTCAACTGCCAGGTGGACCTGGCCTCT 270

Qy     181 CAGGTGCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     271 CAGGTCCCTTGTTTGGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTGC 330

Qy     241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300
        || | || ||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db     331 CAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAAGTGGGCAGATG 390

Qy     301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360
        ||||| ||||| ||||| || ||||| ||||| || || || ||||| ||||| |||||
Db     391 CTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGC 450

Qy     361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420
        |||| ||||| ||||| || ||||| || || ||||| ||||| ||||| |||||
Db     451 CGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCAGCTCG 510

Qy     421 CCTCAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCAAC 480
        ||||| ||||| ||||| || || || ||||| ||||| ||||| ||||| |||||
Db     511 CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACACCAGCTGCTCCCAAC 570

Qy     481 CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC 540
        |||| || ||||| |||| ||||| ||||| ||||| ||||| ||||| |||||
Db     571 TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCC 630

```

Qy	541	CAGGCCAGCGTGACAAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	631		
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	660
Db	691		
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Db	751		
Qy	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAG	780
Db	811		
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	871		
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	931		
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCTCGCTATAGCAACCCTGCGGACTTC	960
Db	991		
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCGAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Db	1051		
Qy	1021	AAGGCACAGTCTCTTGCAGCCCTGTTCTTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1111		
Qy	1081	TGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTGACCCCTCACA	1140
Db	1171		
Qy	1141	CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCCGGATGATAGAGCAGTTTTCC	1197
Db	1231		
Qy	1198	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCAT	1257
Db	1291		
Qy	1258	GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1317
Db	1351		
Qy	1318	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1377
Db	1411		
Qy	1378	TTCAATGTCATCCTGGATGTCGTCCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT	1437

Db	1471	TTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT	1530
Qy	1438	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTCTTTGCCAAGATCCTAGGAGAA	1497
Db	1531	GAAGCTGGAAGACGGGCTGTACACCACTGGTCCATATTCTTTGCCAAGATCCTCGGGGAG	1590
Qy	1498	TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1557
Db	1591	CTTCCGGAGCACTGTGCCTACATCATCTACGGGATGCCACCTACTGGCTGGCCAAC	1650
Qy	1558	CTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC	1617
Db	1651	CTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC	1710
Qy	1618	TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCTCTCC	1677
Db	1711	TGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCACTTCCACATGGCCTCC	1770
Qy	1678	TTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG	1737
Db	1771	TTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG	1830
Qy	1738	GACAACTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTC	1797
Db	1831	AGCAGCCTGTGGACAGTGCCCGCTGGATTTCCAAAGTGCCTTCCTGCGGTGGTGTTTT	1890
Qy	1798	TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC	1857
Db	1891	GAAGGGCTGATGAAGATTTCAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTC	1950
Qy	1858	ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC	1917
Db	1951	ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTC	2010
Qy	1918	TATGCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA	1977
Db	2011	TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCTGTACTACGTG	2070
Qy	1978	TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019
Db	2071	TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA	2112

RESULT 9

AX478099

LOCUS AX478099 3239 bp DNA linear PAT 12-AUG-2002

DEFINITION Sequence 29 from Patent WO0240541.

ACCESSION AX478099

VERSION AX478099.1 GI:22217059

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Tang,Y.T., Yue,H., Nguyen,D.B., Hafalia,A.J., Elliott,V.S., Lu,Y.,

TITLE	Transporters and ion channels
JOURNAL	Patent: WO 0240541-A 29 23-MAY-2002; Incyte Genomics, Inc. (US)

ORIGIN

Qy	884	GGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCTCGCTATA	943
Db	12	GGGCGGCGCCAGCACATGGTCCATTATTTACAGCCATCGGCTACCCCTGTCTCTCGCTACA	71
Qy	944	GCAACCCTGCGGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGG	1003
Db	72	GCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGG	131
Qy	1004	AGGTGGCCACCGTGGAGAAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAG	1063
Db	132	AATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCGTG	191
Qy	1064	GCTTTGATGACTTTTCTGTGGAAGCTGAGGCAAAGGAAGTCAACACAAGCACCCACACAG	1123
Db	192	ACTTAGATGACTTTTCTATGGAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGG	251
Qy	1124	TCAGCCTGACCCTCACACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGA	1180
Db	252	AAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGG	311
Qy	1181	TGATAGAGCAGTTTTTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGC	1240
Db	312	CGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGC	371
Qy	1241	CCACGCTGCTCATTCATGGGTGCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTT	1300
Db	372	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	431
Qy	1301	ACTACGGCCATGGGGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGA	1360
Db	432	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA	491
Qy	1361	TAGGGGCGCTCATTCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGA	1420
Db	492	TCGGTGCTCTCATCCCTTTCAACGTCAATTCTGGATGTCATCTCCAAATGTTACTCAGAGA	551
Qy	1421	GGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTG	1480

TITLE Direct Submission
 JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 204584)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 204584)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 COMMENT On Nov 4, 2003 this sequence version replaced gi:34495085.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: M_BA0148C10

FEATURES Location/Qualifiers
 source 1. .204584
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="17"
 /clone="RP23-148C10"

ORIGIN

Query Match 15.0%; Score 302.2; DB 10; Length 204584;
 Best Local Similarity 80.1%; Pred. No. 9.4e-59;
 Matches 411; Conservative 0; Mismatches 8; Indels 94; Gaps 1;

Qy 559 CGGGTGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGC 618
 | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 204178 CAGGTGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGC
 204119

Qy 619 AACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTTGGGGTGCAG 678
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 204118 AACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTTGGGGTGCAG
 204059

Qy 679 CTCCTGTGGAA----- 689
 |||||||||
 Db 204058 CTCCTGTGGAAACCCAGGTGAGGCCTGGGAACCTGAGGGGTGAAAACCTGAGCCTACAACC
 203999

Qy 690 -----CCCAGGAATCCTCAT 704
 |||||||||||
 Db 203998 TGTCCGGCAGCGGCAGCGTGGTCATTGGACTCCCTGTGCAATATCCCCAGGAATCCTCAT
 203939

Qy 705 TCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTT 764

```

|||||
Db      203938 TCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTT
203879

Qy      765 GTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGA 824
|||||

Db      203878 GTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGA
203819

Qy      825 CATCTTCAGGCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGG 884
|||||

Db      203818 CATCTTCAGGCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGG
203759

Qy      885 GGCGGCGCAGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCTCGCTATAG 944
|||||

Db      203758 GGCGGCGCAGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCTCGCTATAG
203699

Qy      945 CAACCTGCGGACTTCTACGTGGACTTGACCAG 977
|||||

Db      203698 CAACCTGCGGACTTCTACGGTGAGTGGTAAAG 203666

```

RESULT 11

F351799S06

LOCUS F351799S06 1387 bp DNA linear ROD 23-AUG-2002

DEFINITION Mus musculus sterolin 2 (Abcg8) gene, exon 6.

ACCESSION AF351804

VERSION AF351804.1 GI:18996442

KEYWORDS .

SEGMENT 6 of 13

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1387)

AUTHORS Lu,K., Lee,M.-H., Yu,H., Zhou,Y., Sandell,S.A., Salen,G. and
Patel,S.B.

TITLE Molecular cloning, genomic organization, genetic variations, and
characterization of murine sterolin genes Abcg5 and Abcg8

JOURNAL J. Lipid Res. 43 (4), 565-578 (2002)

MEDLINE 21904563

PUBMED 11907139

REFERENCE 2 (bases 1 to 1387)

AUTHORS Lu,K., Zhou,Y., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St., STB 541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers

source 1..1387

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="129/Sv"

/db_xref="taxon:10090"

/chromosome="17"

St., STB 541, Charleston, SC 29403, USA

FEATURES

source 1. .1378

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="129/Sv"

/db_xref="taxon:10090"

/chromosome="17"

/map="between Mit41 and Mit189"

/clone="329B11"

exon 415. .682

/gene="Abcg8"

/number=11

ORIGIN

Query Match 13.4%; Score 270.6; DB 10; Length 1378;

Best Local Similarity 88.3%; Pred. No. 1.5e-51;

Matches 294; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1484 AGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCT 1543

|||||

Db 413 AGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCT 472

Qy 1544 ACTGGCTGACAAACCTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGT 1603

|||||

Db 473 ACTGGCTGACAAACCTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGT 532

Qy 1604 GGTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCT 1663

|||||

Db 533 GGTGGTGGTCTTCTGCTGCAGGAACATGGCCCTGGCTGCCTCTGCCATGCTGCCACCT 592

Qy 1664 TCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCT 1723

|||||

Db 593 TCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCT 652

Qy 1724 TCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCC 1783

|||||

Db 653 TCATGATAAACTTGGACAACCTGTGGATAGGTGAGGCCTGCTGCCCCACCCCCGCCCCC 712

Qy 1784 TCCGGTGGTGGTCTTCTCGGGGCTGATGCAGATTC 1816

| || ||| ||| || |||

Db 713 CTTAGCCAAGCGTCTGTAGGCCTCTGTGGCTGC 745

RESULT 13

AC120701

LOCUS AC120701 237445 bp DNA linear HTG 21-SEP-2002

DEFINITION Rattus norvegicus clone CH230-65H6, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.

ACCESSION AC120701

VERSION AC120701.4 GI:23265381

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 237445)

AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hagues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 237445)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 237445)
 AUTHORS Rat Genome Sequencing Consortium.
 TITLE Direct Submission
 JOURNAL Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Sep 21, 2002 this sequence version replaced gi:21908396.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
 sequence may extend beyond the ends of the clone and there may be
 contigs that consist entirely of whole genome shotgun sequence
 reads. Both end sequences and whole genome shotgun sequence only
 contigs will be indicated in the feature table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GXQV
 Center clone name: CH230-65H6
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 209781 bases at least Q40
 Consensus quality: 213033 bases at least Q30
 Consensus quality: 214997 bases at least Q20
 Estimated insert size: 233017; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 233866: contig of 233866 bp in length
 * 233867 233966: gap of unknown length
 * 233967 235011: contig of 1045 bp in length
 * 235012 235111: gap of unknown length
 * 235112 236137: contig of 1026 bp in length
 * 236138 236237: gap of unknown length
 * 236238 237445: contig of 1208 bp in length.
 FEATURES Location/Qualifiers
 source 1. .237445
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-65H6"
 misc_feature 1. .1326
 /note="wgs_end_extension
 clone_end:T7"

```

misc_feature      8065. .8944
                  /note="clone_boundary
                  clone_end:T7
                  site:EcoRI
                  end_sequence:BH350813"
misc_feature      complement(232953. .233569)
                  /note="clone_boundary
                  clone_end:Sp6
                  site:EcoRI
                  end_sequence:BH350815"

```

ORIGIN

```

Query Match          13.1%;  Score 264.8;  DB 2;  Length 237445;
Best Local Similarity 76.3%;  Pred. No. 4e-50;
Matches 380;  Conservative 0;  Mismatches 32;  Indels 86;  Gaps 1;

```

```

Qy      559  CGGGTGGAAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGC 618
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      146962 CAGGTGGAAGACGTGATTGCGGAGCTGCGGCTGCGGCAGTGCGCCAACACCCGCGTGGGC
147021

Qy      619  AACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTTGGGGTGCAG 678
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147022 AACACATACGTACGCGGGGTGTCCGGGGGCGAGCGCCGAAGAGTGAGCATCGGGGTGCAG
147081

Qy      679  CTCCTGTGGAA----- 689
          | | | | | | | | | |
Db      147082 CTCCTGTGGAACCCAGGTGAGGCCTGGGAACCTGAGGGGCGAGGACCTGAGCCTACAACC
147141

Qy      690  -----CCCAGGAATCCTCATCTGGATG 712
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147142 TGTCCGGCGTGGTCACTGGGCTCCCTGTGCGATACCCCCAGGAATCCTCATCTGGATG
147201

Qy      713  AACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCC 772
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147202 AACCCACTTCCGGCCTCGACAGCTTCACCGCTCACAACCTGGTGAGAACTTTGTCCCGCC
147261

Qy      773  TGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCA 832
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147262 TGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCA
147321

Qy      833  GGCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGGCGC 892
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147322 GGCTATTTGACCTGGTCCTTCTGATGACGTCTGGCACCCCTATCTACCTGGGGGTGGCAC
147381

Qy      893  AGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTG 952
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147382 AGCACATGGTGCAGTACTTTACATCAATTGGCTACCCTTGTCCTCGCTACAGCAACCCTG
147441

Qy      953  CGGACTTCTACGTGGACT 970

```

| | | | | | | | | | | | | | |
Db 147442 CTGACTTCTACGGTGAGT 147459

RESULT 14

AC112747/c

LOCUS AC112747 312858 bp DNA linear HTG 08-OCT-2002

DEFINITION Rattus norvegicus clone CH230-359E1, *** SEQUENCING IN PROGRESS
***, 8 unordered pieces.

ACCESSION AC112747

VERSION AC112747.3 GI:23270105

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 312858)

AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hagues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 312858)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 312858)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 23, 2002 this sequence version replaced gi:21738477.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRAX
Center clone name: CH230-359E1
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 241372 bases at least Q40
Consensus quality: 245333 bases at least Q30
Consensus quality: 248022 bases at least Q20
Estimated insert size: 276767; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently


```

                                     |||
Db      82048 TGTCCGGCGTGGTCACTGGGCTCCCTGTGCGATACCCCCAGGAATCCTCATCCTGGATG 81989

Qy      713 AACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCC 772
        |||
Db      81988 AACCCACTTCCGGCCTCGACAGCTTCACCGCTCACAACCTGGTGAGAACTTTGTCCCGCC 81929

Qy      773 TGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCA 832
        |||
Db      81928 TGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCA 81869

Qy      833 GGCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGC 892
        |||
Db      81868 GGCTATTTGACCTGGTCCTTCTGATGACGTCTGGCACCCCTATCTACCTGGGGGTGGCAC 81809

Qy      893 AGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTG 952
        |||
Db      81808 AGCACATGGTGCAGTACTTTACATCAATTGGCTACCCTTGTCTCGCTACAGCAACCCTG 81749

Qy      953 CGGACTTCTACGTGGACT 970
        |||
Db      81748 CTGACTTCTACGGTGAGT 81731

```

RESULT 15

AY145899/c

LOCUS AY145899 40929 bp DNA linear ROD 12-NOV-2002

DEFINITION Rattus norvegicus sterolin 2 (Abcg8) and sterolin 1 (Abcg5) genes, complete cds.

ACCESSION AY145899

VERSION AY145899.1 GI:24935208

KEYWORDS .

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 40929)

AUTHORS Yu,H., Lu,K., Lee,M., Pandit,B. and Patel,s.B.

TITLE The rat Abcg5 and Abcg8: characterization, chromosomal assignment and genetic variation in sitosterolemic rats

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 40929)

AUTHORS Yu,H., Lu,K., Lee,M., Pandit,B. and Patel,s.B.

TITLE Direct Submission

JOURNAL Submitted (29-AUG-2002) Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty Street, STR 541, Charleston, SC 29403, USA

FEATURES Location/Qualifiers

source 1. .40929

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

gene complement(<4136. .>20831)

/gene="Abcg8"

mRNA complement(join(<4136. .4273,4361. .4488,5693. .5960,

6513. .6589,6754. .6953,8189. .8269,8350. .8512,10772.

.11041,

11129. .11261,11647. .11885,15513. .15669,17473. .17574,
20769. .>20831))
/gene="Abcg8"
/product="sterolin 2"
CDS complement(join(4136. .4273,4361. .4488,5693. .5960,
6513. .6589,6754. .6953,8189. .8269,8350. .8512,10772.

.11041,

11129. .11261,11647. .11885,15513. .15669,17473. .17574,
20769. .20831))
/gene="Abcg8"
/note="ATP-binding cassette sub-family G (WHITE) member 8"
/codon_start=1
/product="sterolin 2"
/protein_id="AAN64276.1"
/db_xref="GI:24935210"
/translation="MAEKTKEETQLWNGTVLQDASSLQDSVFSSSDNSLYFTYSGQS
NTLEVRDLTYQVDMASQVPWFQQLAQFKLPWRSRGSQDSWDLGIRNLSFKVRSQGMLA
IIGSAGCGRATLLDVITGRDHGGKMKSGQIWINQGPSTPQLIQKCVAHVRQQDQLLPN
LTVRETTLTFIAQMRLPKTFSAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER
RRVSIGVQLLWNPGLILDEPTSGLDSEFTAHNLVRTLSRLAKGNRLVLISLHQPRSDI
FRLFDLVLLMTSGTPIYLGVAQHMVQYFTSIGYPCPRYSNPADFYVDLTSIDRRSKEQ
EVATMEKARLLAALFLEKVQGFDDFLWKAELSLDTGTAYVSQTLTQDTNCGTAAELP
GMIQQFTTLIRRQISNDFRDLPTLFIHGAEACLSLIIGFLYYGHADKPLSFMDMAAL
LFMIGALIPFNVILDVVSCHSERSLLYYELEDGLYTAGPYFFAKVLGELPEHCAYVI
IYGMPIYWLTNLRPGPELFLHFMLLWLVVFCCRTMALAASAMLPFHMSSFCCNALY
NSFYLTAGFMINLNNLWIVPAWISKMSFLRWCFSGLMQIQFNHIIYTTQIGNLTFSPV
GDAMVTAMDLSHPLYAIYLIVIGISCGFLSLYYLSLKFIKQKSIQDW"
gene <21211. .>40564
/gene="Abcg5"
mRNA join(<21211. .21356,21968. .22089,24726. .24862,24949.
.25047,

27388. .27520,28838. .28977,29879. .30008,30715. .30928,
31032. .31237,32869. .33007,35821. .36006,38553. .38665,
40371. .>40564)
/gene="Abcg5"
/product="sterolin 1"
CDS join(21211. .21356,21968. .22089,24726. .24862,24949.
.25047,

27388. .27520,28838. .28977,29879. .30008,30715. .30928,
31032. .31237,32869. .33007,35821. .36006,38553. .38665,
40371. .40564)
/gene="Abcg5"
/note="ATP-binding cassette sub-family G (WHITE) member 5"
/codon_start=1
/product="sterolin 1"
/protein_id="AAN64275.1"
/db_xref="GI:24935209"
/translation="MSELPFLSPEGARGPHNNRGSQSSLEEGSVTGSEARHSLGVLNV
SFSVSNRVGPWWNIKSCQKQWDRKILKDVSLYIESGQTMCILGSSSGSKTTLLDAISG
RLRRTGTLEGEVFNVCCELRRDQFQDCVSYLLQSDVFLSSLTVRETTLRYTAMLALRSS
SADFYDKKVEAVLTLSLHVADQMIGNYNFGGISSGERRRVSIAAQLLQDPKVMMLD
EPTTGLDCMTANHIVLLVELARRNRIVIVTIHQPRSELFHHFDKIAILTYGELVFCEG
TPEEMLGFFNCCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLESAFRQSD
ICHKILENIERTRHLKTLPMVPFKTKNPPGMFCKLGVLRLRVTRNLMRNKQVVIMRLV
QNLIMGLFLIFYLLRVQNNMLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVS

DQESQDGLYQKWQMLLAYVLHALPFSIVATVIFSSVCYWTGLGLYPEVARFGYFSAALL
APHLIGEFLLVLLGMVQNPNIIVNSIVALLSISGLLIGSGFIRNIEEMPIPLKILGYF
TFQKYCCEILVVNEFYGLNFTCGGSNTSVPNPMCSMTQGIQFIEKTCPGATSRFTTN
FLILYSFIPTLVILGMVVFKVRDYLISR"

ORIGIN

Query Match 13.1%; Score 264.2; DB 10; Length 40929;
Best Local Similarity 68.2%; Pred. No. 5.2e-50;
Matches 429; Conservative 0; Mismatches 113; Indels 87; Gaps 1;

```
Qy      429 GGTGAGGAAGTGCCTTGCATGTGCGGCAGCATGACCAACTGCTGCCAACCTGACCGT 488
      ||| || | | | | | | | | | | | | | | | | | | | |
Db     11394 GGTGGGGGTGGGGGTGGGGGTGGGGGAGTTCGCCAAACAATGCTGCAGGAAATGAAGTG 11335

Qy      489 CAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCCA 548
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     11334 CAGAAGAAGCCTTCCTTGGCATTAAAGTGAAGAAGTTGCCCCCTGGACGCTCGTAATGCTC 11275

Qy      549 GCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACAC 608
      | | | | | | | | | | | | | | | | | | | | | | | |
Db     11274 AGCCTGCCCTCAGGTGGAAGACGTGATTGCGGAGCTGCGGCTGCGGCAGTGCGCCAACAC 11215

Qy      609 CAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCAT 668
      | | | | | | | | | | | | | | | | | | | | | | | |
Db     11214 CCGCGTGGGCAACACATACGTACGCGGGGTGTCCGGGGGCGAGCGCCGAAGAGTGAGCAT 11155

Qy      669 TGGGGTGCAGCTCCTGTGGA----- 689
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     11154 CGGGGTGCAGCTCCTGTGGAACCCAGGTGAGGCCTGGGAACCTGAGGGGCGAGGACCTGA 11095

Qy      690 -----CCAGGAATCCT 701
      | | | | | | | | | | | | | | | | | | | | | | | |
Db     11094 GCCTACAACCTGTCCGGCGTGGTCACTGGGCTTCCCTGTGCGATACCCCCAGGAATCCT 11035

Qy      702 CATCTTGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAAC 761
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     11034 CATCCTGGATGAACCCACTTCCGGCCTCGACAGCTTCACCGCTCACAACCTGGTGAGAAC 10975

Qy      762 CTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTC 821
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     10974 TTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTC 10915

Qy      822 TGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCT 881
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     10914 TGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGACGCTCTGGCACCCCTATCTACCT 10855

Qy      882 GGGGGCGGCGCAGCAAATGGTGCACTACTTCACATCCATTGGCCACCCTTGTCTCTCGCTA 941
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     10854 GGGGGTGGCACAGCACATGGTGCACTACTTTACATCAATTGGCTACCCTTGTCTCTCGCTA 10795

Qy      942 TAGCAACCCTGCGGACTTCTACGTGGACT 970
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     10794 CAGCAACCCTGCTGACTTCTACGGTGAGT 10766
```

Search completed: February 26, 2004, 06:21:11
Job time : 5199.97 secs

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:39:18 ; Search time 511.357 Seconds
(without alignments)
16773.223 Million cell updates/sec

Title: US-09-989-981A-3
Perfect score: 2019
Sequence: 1 atggctgagaaaaccaaaga.....agtcaattcaagactggtga 2019

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2019	100.0	2019	7	AAD48881	Aad48881 Mouse ABC
2	2004.4	99.3	2564	6	ABN90022	Abn90022 Mouse clo
3	1430	70.8	2669	7	AAD48883	Aad48883 Human ABC
4	743.8	36.8	3239	6	ABK83218	Abk83218 Human tra
5	229.2	11.4	580	4	AAH98911	Aah98911 Arabidops
6	199.2	9.9	1920	6	ABK51681	Abk51681 DNA encod
7	199.2	9.9	2340	6	AAD22009	Aad22009 Human sit

8	199.2	9.9	2340	7	AAD48882	Aad48882	Human	ABC
9	199.2	9.9	2516	6	ABK51682	Abk51682	Human	ABC
10	195	9.7	2035	6	ABK51686	Abk51686	cDNA	enco
11	186.6	9.2	1915	6	ABK51684	Abk51684	DNA	encod
12	186.6	9.2	1959	7	AAD48880	Aad48880	Mouse	ABC
13	186.6	9.2	2258	6	AAD22008	Aad22008	Mouse	sit
14	186.6	9.2	2354	6	ABK51685	Abk51685	Mouse	ABC
15	176	8.7	1069	6	ABK51687	Abk51687	cDNA	enco
16	137.8	6.8	2525	3	AAZ98625	Aaz98625	Silkworm	
17	137	6.8	2025	6	ABA94371	Aba94371	Murine	BC
18	133.6	6.6	363	6	ABN16253	Abn16253	Human	ORF
19	132.4	6.6	1968	6	AAL42412	Aal42412	Human	BCR
20	132.4	6.6	1968	9	ADC54181	Adc54181	Human	bre
21	132.4	6.6	1998	6	AAL42414	Aal42414	Human	BCR
22	132.4	6.6	2027	6	ABK49901	Abk49901	cDNA	enco
23	132.4	6.6	2053	6	ABK49911	Abk49911	cDNA	enco
24	132.4	6.6	2247	6	ABA94383	Aba94383	Human	BCR
25	132.4	6.6	2418	2	AAZ06360	Aaz06360	Breast	Ca
26	132.4	6.6	2574	4	AAF27724	Aaf27724	Human	tra
27	132.4	6.6	2574	8	ADA10916	Adal0916	Human	cDN
28	132.4	6.6	2718	7	ACC80605	Acc80605	Human	ABC
29	132.4	6.6	2719	3	AAZ94760	Aaz94760	Human	ATP
30	132.4	6.6	2719	3	AAA27938	Aaa27938	ATP-bindi	
31	132.4	6.6	2719	6	ABA94369	Aba94369	Human	BCR
32	132.4	6.6	2883	6	ABZ35528	Abz35528	Human	gen
33	130.8	6.5	1998	6	AAL42413	Aal42413	Human	BCR
34	125.8	6.2	2446	3	AAC37975	Aac37975	Arabidops	
35	119.4	5.9	727	4	AAH07859	Aah07859	Human	cDN
36	119.4	5.9	2077	4	AAH15008	Aah15008	Human	cDN
37	117.8	5.8	2352	4	ABL05135	Abl05135	Drosophil	
38	117.2	5.8	2930	3	AAZ94747	Aaz94747	Human	ATP
39	117.2	5.8	2930	6	ABL63321	Abl63321	Breast	ca
40	117	5.8	3201	6	ABV74352	Abv74352	Human	ABC
41	114	5.6	2133	9	ADE47651	Ade47651	Human	NOV
42	107	5.3	2921	7	ABV75074	Abv75074	Human	Dev
43	106.4	5.3	4646	7	ADA68676	Ada68676	Spirodela	
44	105.4	5.2	2894	7	ACD13444	Acd13444	Human	DNA
c 45	105	5.2	6043	7	AAD48884	Aad48884	ABCG5-ABC	

ALIGNMENTS

RESULT 1

AAD48881

ID AAD48881 standard; DNA; 2019 BP.

XX

AC AAD48881;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG8 DNA.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5; gene; ds.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1. .2019
 FT /*tag= a
 FT /product= "mABCG8 protein"
 FT /transl_except= (pos:1318. .1320, aa:Leu)
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR P-PSDB; AAE31703.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 13; Page 75; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is mouse ABCG8 DNA
 XX
 SQ Sequence 2019 BP; 444 A; 598 C; 510 G; 467 T; 0 U; 0 Other;

Query Match 100.0%; Score 2019; DB 7; Length 2019;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT	60
Db	1	ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT	60
Qy	61	TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC	120
Db	61	TCGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC	120

Qy 121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
 |||
 Db 121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180

Qy 181 CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
 |||
 Db 181 CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240

Qy 241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300
 |||
 Db 241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300

Qy 301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360
 |||
 Db 301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360

Qy 361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420
 |||
 Db 361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420

Qy 421 CCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCCAAC 480
 |||
 Db 421 CCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCCAAC 480

Qy 481 CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCAGATGCGCCTGCCAGGACCTTCTCC 540
 |||
 Db 481 CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCAGATGCGCCTGCCAGGACCTTCTCC 540

Qy 541 CAGGCCACGCTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC 600
 |||
 Db 541 CAGGCCACGCTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC 600

Qy 601 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCGGGGGTGAGCGCCGACGA 660
 |||
 Db 601 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCGGGGGTGAGCGCCGACGA 660

Qy 661 GTGAGCATTGGGGTGACAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 720
 |||
 Db 661 GTGAGCATTGGGGTGACAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 720

Qy 721 TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGGCCTGGCCAAG 780
 |||
 Db 721 TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGGCCTGGCCAAG 780

Qy 781 GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT 840
 |||
 Db 781 GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT 840

Qy 841 GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG 900
 |||
 Db 841 GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG 900

Qy 901 GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC 960
 |||
 Db 901 GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC 960

Qy 961 TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG 1020

Db	961		TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Qy	1021		AAGGCACAGTCTCTTGCAGCCCTGTTCC TAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1021		AAGGCACAGTCTCTTGCAGCCCTGTTCC TAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Qy	1081		TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1081		TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Qy	1141		CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCCACC	1200
Db	1141		CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTTCCACC	1200
Qy	1201		CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATT CATGGG	1260
Db	1201		CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATT CATGGG	1260
Qy	1261		TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Db	1261		TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Qy	1321		CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Db	1321		CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Qy	1381		AATGTCATCCTGGATGTCTCTCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1440
Db	1381		AATGTCATCCTGGATGTCTCTCAAATGTCACTCGGAGAGGTCAATGCTGTACTATGAG	1440
Qy	1441		CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Db	1441		CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Qy	1501		CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Db	1501		CCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Qy	1561		CGGCCCCGTGCCTGAGCTCTTCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620
Db	1561		CGGCCCCGTGCCTGAGCTCTTCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620
Qy	1621		TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCCTCCTTC	1680
Db	1621		TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCCTCCTTC	1680
Qy	1681		TTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC	1740
Db	1681		TTCTGCAATGCCCTCTACAACCTCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC	1740
Qy	1741		AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTCTCG	1800
Db	1741		AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTCTCG	1800
Qy	1801		GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC	1860

Db 1801 GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC 1860

Qy 1861 TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT 1920
 |||

Db 1861 TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT 1920

Qy 1921 GCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTTCCTGTACTATCTATCC 1980
 |||

Db 1921 GCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTTCCTGTACTATCTATCC 1980

Qy 1981 TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
 |||

Db 1981 TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019

RESULT 2

ABN90022

ID ABN90022 standard; cDNA; 2564 BP.

XX

AC ABN90022;

XX

DT 16-AUG-2002 (first entry)

XX

DE Mouse clone IMX3_67 extended sequence.

XX

KW Mouse; antiinflammatory; gene therapy; ileitis; DST; ss; TOGA;

KW digital sequence tag; total gene expression analysis.

XX

OS Mus musculus.

XX

PN WO200231114-A2.

XX

PD 18-APR-2002.

XX

PF 11-OCT-2001; 2001WO-US032091.

XX

PR 11-OCT-2000; 2000US-0239483P.

XX

PA (DIGI-) DIGITAL GENE TECHNOLOGIES INC.

XX

PI Viney JL, Sims JE, Dubose RF, Baum PR, Hasel KW, Hilbush BS;

XX

DR WPI; 2002-426279/45.

XX

PT New isolated nucleic acid molecules that are associated with ileitis, for

PT preventing, treating, modulating and diagnosing ileitis in a mammalian

PT subject.

XX

PS Claim 1; Page 266-268; 273pp; English.

XX

CC The invention relates to a novel isolated nucleic acid molecule
 CC comprising a polynucleotide having one of 90 polynucleotide sequences,
 CC given in the specification. The polynucleotides of the invention have
 CC antiinflammatory activity, and may have a use in gene therapy. The
 CC polynucleotide or a polypeptide encoded by it is used for preventing,
 CC treating, modulating or ameliorating a medical condition such as ileitis.
 CC The polypeptide or polynucleotide is also useful for manufacturing a

CC medicament for treating ileitis. The sequence represents a an extended
CC cDNA digital sequence tag obtained from a mouse clone by the TOGA (total
CC gene expression analysis) method
XX
SQ Sequence 2564 BP; 623 A; 722 C; 638 G; 581 T; 0 U; 0 Other;

Query Match 99.3%; Score 2004.4; DB 6; Length 2564;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2018; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

```
Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
          |||
Db      35 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 94

Qy      61 TC---GGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 117
          ||
Db      95 TCGCAGGGCCTCCAGGACAGCTTGTTCTCCTCGGAAAGTGACAACAGTCTGTACTTCACC 154

Qy     118 TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC 177
          |||
Db     155 TACAGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCC 214

Qy     178 TCTCAGGTGCCTTGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 237
          |||
Db     215 TCTCAGGTGCCTTGTTTGTAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGC 274

Qy     238 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 297
          |||
Db     275 AGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAG 334

Qy     298 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 357
          |||
Db     335 ATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACA 394

Qy     358 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 417
          |||
Db     395 GGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGT 454

Qy     418 ACGCCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCC 477
          |||
Db     455 ACGCCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCC 514

Qy     478 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTC 537
          |||
Db     515 AACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTC 574

Qy     538 TCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAG 597
          |||
Db     575 TCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAG 634

Qy     598 TGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGA 657
          |||
Db     635 TGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGA 694

Qy     658 CGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCC 717
          |||
Db     695 CGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCC 754
```

Qy	718	ACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCC	777
Db	755	ACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCC	814
Qy	778	AAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTA	837
Db	815	AAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTA	874
Qy	838	TTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAA	897
Db	875	TTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAA	934
Qy	898	ATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGAC	957
Db	935	ATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGAC	994
Qy	958	TTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTG	1017
Db	995	TTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTG	1054
Qy	1018	GAGAAGGCACAGTCTCTTGCAGCCCTGTTTCCTAGAAAAAGTACAAGGCTTTGATGACTTT	1077
Db	1055	GAGAAGGCACAGTCTCTTGCAGCCCTGTTTCCTAGAAAAAGTACAAGGCTTTGATGACTTT	1114
Qy	1078	CTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTC	1137
Db	1115	CTGTGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTC	1174
Qy	1138	ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTTCC	1197
Db	1175	ACACAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTTCC	1234
Qy	1198	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCAT	1257
Db	1235	ACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCAT	1294
Qy	1258	GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1317
Db	1295	GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1354
Qy	1318	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1377
Db	1355	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1414
Qy	1378	TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT	1437
Db	1415	TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT	1474
Qy	1438	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA	1497
Db	1475	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA	1534
Qy	1498	TTGCCGGAGCACTGTGCCTACGTATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1557
Db	1535	TTGCCGGAGCACTGTGCCTACGTATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1594

Qy 1558 CTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC 1617
 |||
 Db 1595 CTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC 1654
 Qy 1618 TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCTCTCC 1677
 |||
 Db 1655 TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCTCTCC 1714
 Qy 1678 TTCTTCTGCAATGCCCTCTACAACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG 1737
 |||
 Db 1715 TTCTTCTGCAATGCCCTCTACAACTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG 1774
 Qy 1738 GACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTC 1797
 |||
 Db 1775 GACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTC 1834
 Qy 1798 TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC 1857
 |||
 Db 1835 TCGGTGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC 1894
 Qy 1858 ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC 1917
 |||
 Db 1895 ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC 1954
 Qy 1918 TATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA 1977
 |||
 Db 1955 TATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA 2014
 Qy 1978 TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
 |||
 Db 2015 TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2056

RESULT 3

AAD48883

ID AAD48883 standard; DNA; 2669 BP.

XX

AC AAD48883;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG8 DNA.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 100..2121

FT /*tag= a

FT /product= "hABCG8 protein"

XX

PN WO200281691-A2.

XX PD 17-OCT-2002.
XX
PF 20-NOV-2001; 2001WO-US043823.
XX
PR 20-NOV-2000; 2000US-0252235P.
PR 28-NOV-2000; 2000US-0253645P.
XX
PA (TULA-) TULARIK INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tian H;
XX
DR WPI; 2003-058548/05.
DR P-PSDB; AAE31705.
XX
PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
PT related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies.
XX
PS Claim 13; Page 80; 94pp; English.
XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is human ABCG8 DNA
XX
SQ Sequence 2669 BP; 595 A; 768 C; 722 G; 584 T; 0 U; 0 Other;

Qy	1	ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT	60
Db	100	ATGGCCGGGAAGGCGGCAGAGGAGAGAGGGCTGCCGAAAGGGGCCACTCCCCAGGATACC	159
Qy	61	TCGGGCCTCCAGGACAGCTTGTTCCTCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC	120
Db	160	TCGGGCCTCCAGGATAGATTGTTCTCTCTGAAAGTGACAACAGCTGTACTTCACCTAC	219
Qy	121	AGTGGTCAGTCCAACACTCTGGAGGTCTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT	180
Db	220	AGTGGCCAGCCCAACACCCTGGAGGTCTCAGAGACCTCAACTACCAGGTGGACCTGGCCTCT	279
Qy	181	CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC	240
Db	280	CAGGTCCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATGCCCTGGACATCTCCCAGCTGC	339
Qy	241	CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG	300
Db	340	CAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGCAGATG	399

Qy	301	CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGG	360
Db	400	CTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCACTGGC	459
Qy	361	AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG	420
Db	460	CGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCAGCTCG	519
Qy	421	CCTCAGCTGGTGAGGAAGTGCCTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCAAC	480
Db	520	CCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAACCAGCTGCTCCCCAAC	579
Qy	481	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCC	540
Db	580	TTGACTGTGCGAGAGACCTTGGCCTTCATTGCCCAGATGCGGCTGCCCAGAACCTTCTCC	639
Qy	541	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	640	CAGGCCCAGCGTGACAAAAGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGCAGTGC	699
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	660
Db	700	GCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGTGTGAGCGCAGGAGA	759
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Db	760	GTCAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTTATTCTCGACGAACCCACC	819
Qy	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAG	780
Db	820	TCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTTGTCCAGGCTGGCCAAA	879
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	880	GGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTGTTT	939
Qy	841	GACCTGGTCTCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	940	GATCTGGTCTCTCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCCAGCACATG	999
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCCTGCGGACTTC	960
Db	1000	GTCCAGTATTTACAGCCATCGGCTACCCCTGTCTCGCTACAGCAATCCTGCTGACTTC	1059
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Db	1060	TATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCGGGAG	1119
Qy	1021	AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1120	AAGGCTCAGTCACTCGAGCCCTGTTTCTAGAAAAAGTGC GTGACTTAGATGACTTTCTA	1179
Qy	1081	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1180	TGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGGAAAGCAGCGTGACCCCA	1239

Qy	1141	CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCGGGATGATAGAGCAGTTTCC	1197
Db	1240	CTAGACACCAACTGCCTCCCGAGTCTACGAAGATGCCTGGGGCGGTGCAGCAGTTTACG	1299
Qy	1198	ACCGTGATCCGTGCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCAT	1257
Db	1300	ACGCTGATCCGTGCGTCAGATTTCCAACGACTTCCGAGACCTGCCACCCTCCTCATCCAT	1359
Qy	1258	GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC	1317
Db	1360	GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTGGCCATGGGAGC	1419
Qy	1318	AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT	1377
Db	1420	ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGATCGGTGCTCTCATCCCT	1479
Qy	1378	TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT	1437
Db	1480	TTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT	1539
Qy	1438	GAGCTGGAAGACGGGCTGTACACTGCTGGTCCCTATTTCTTTGCCAAGATCCTAGGAGAA	1497
Db	1540	GAACCTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG	1599
Qy	1498	TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC	1557
Db	1600	CTTCCGGAGCACTGTGCCTACATCATCATCTACGGGATGCCACCTACTGGCTGGCCAAC	1659
Qy	1558	CTGCGGCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC	1617
Db	1660	CTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC	1719
Qy	1618	TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCTCTCC	1677
Db	1720	TGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCC	1779
Qy	1678	TTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG	1737
Db	1780	TTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG	1839
Qy	1738	GACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCTCCGGTGGTGCTTC	1797
Db	1840	AGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCTTCTGCGGTGGTGTTTT	1899
Qy	1798	TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC	1857
Db	1900	GAAGGGCTGATGAAGATTCAAGTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTC	1959
Qy	1858	ACCTTCTCCATCTCGGAGACACGATGATCAGTGCCATGGACCTGAACCTCGCATCCACTC	1917
Db	1960	ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTC	2019
Qy	1918	TATGCGATCTACCTCATTGTTCATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA	1977
Db	2020	TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTG	2079
Qy	1978	TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019

||||| | | ||||| ||||| ||| ||||| |||||
Db 2080 TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA 2121

RESULT 4

ABK83218

ID ABK83218 standard; cDNA; 3239 BP.

XX

AC ABK83218;

XX

DT 27-AUG-2002 (first entry)

XX

DE Human transporter and ion channel, TRICH9, Incyte ID 6585710CB1, cDNA.

XX

KW Human; ss; gene; transporter and ion channel; TRICH; transport disorder;

KW neurological disorder; muscle disorder; immunological disorder; cancer;

KW scleroderma; systemic lupus erythematosus; allergy; leukaemia;

KW cell proliferative disorder; cervical cancer; breast cancer;

KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;

KW myotonic dystrophy; catatonia; endocrine disorder; diabetes;

KW Grave's disease; gastrointestinal disorder; Crohn's disease;

KW renal disorder; Good pasture's syndrome; viral infection; cirrhosis;

KW bacterial infection; fungal infection; parasitic infection;

KW protozoal infection; helminthic infection; cardiovascular disorder;

KW atherosclerosis; hepatic disease.

XX

OS Homo sapiens.

XX

PN WO200240541-A2.

XX

PD 23-MAY-2002.

XX

PF 25-OCT-2001; 2001WO-US046055.

XX

PR 27-OCT-2000; 2000US-0243989P.

PR 03-NOV-2000; 2000US-0245904P.

PR 09-NOV-2000; 2000US-0247673P.

PR 17-NOV-2000; 2000US-0249661P.

PR 20-NOV-2000; 2000US-0252232P.

PR 01-DEC-2000; 2000US-0250790P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Tang YT, Yue H, Nguyen DB, Hafalia AJA, Elliott VS, Lu Y;

PI Walia NK, Yao MG, Baughn MR, Gandhi AR, Ding L, Sanjanwala M;

PI Ramkumar J, Arvizu C, Gietzen KJ, Lal PG, Azimzai Y, Khan FA;

PI Thangavelu K, Thornton M, Lu DAM, Tribouley CM, Warren BA, Ison CH;

PI Das D, Raumann BE, Policky JL, Kearney L;

XX

DR WPI; 2002-463570/49.

DR P-PSDB; ABG61539.

XX

PT New transporters and ion channels (TRICH) polypeptides, useful for

PT diagnosing, preventing, and treating disorders associated with an

PT abnormal expression or activity of TRICH, e.g. immunological, muscular or

PT renal disorders.

XX

XX

XX

Dlc

Dk

Dk

Dk

Dk

Qy	1181	TGATAGAGCAGTTTTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGC	1240
Db	312	CGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGC	371
Qy	1241	CCACGCTGCTCATTCATGGGTGCGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTT	1300
Db	372	CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT	431
Qy	1301	ACTACGGCCATGGGGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGA	1360
Db	432	ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTGTTCATGA	491
Qy	1361	TAGGGGCGCTCATTCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGA	1420
Db	492	TCGGTGCTCTCATCCCTTTCAACGTCATTCCTGGATGTCATCTCCAAATGTTACTCAGAGA	551
Qy	1421	GGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTG	1480
Db	552	GGGCAATGCTTTACTATGAAGTGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG	611
Qy	1481	CCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCA	1540
Db	612	CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCATCTACGGGATGCCCA	671
Qy	1541	TCTACTGGCTGACAAACCTGCGGCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCG	1600
Db	672	CCTACTGGCTGGCCAACCTGAGGCCAGGCCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGG	731
Qy	1601	TGTGGTTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCA	1660
Db	732	TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA	791
Qy	1661	CCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCG	1720
Db	792	CCTTCCACATGGCCTCCTTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGG	851
Qy	1721	GCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGT	1780
Db	852	GCTTCATGATAAACTTGGACAGCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCCT	911
Qy	1781	TCCTCCGGTGGTGCTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCA	1840
Db	912	TCCTGCGGTGGTGTTTTGAAGGGCTGATGAAGATTAGTTTTCAGCAGAAGAAGTATAAAA	971
Qy	1841	CACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACC	1900
Db	972	TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC	1031
Qy	1901	TGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCC	1960
Db	1032	TGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA	1091
Qy	1961	TGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019
Db	1092	TGGTCTGTACTACGTGTCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA	1150

RESULT 5

AAH98911

ID AAH98911 standard; cDNA; 580 BP.

XX

AC AAH98911;

XX

DT 12-OCT-2001 (first entry)

XX

DE Arabidopsis EST-derived coding sequence SEQ ID NO: 768.

XX

KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;

KW gene therapy; nutrition; ss.

XX

OS Arabidopsis thaliana.

XX

PN WO200154477-A2.

XX

PD 02-AUG-2001.

XX

PF 25-JAN-2001; 2001WO-US002687.

XX

PR 25-JAN-2000; 2000US-00491404.

PR 17-JUL-2000; 2000US-00617746.

PR 03-AUG-2000; 2000US-00631451.

PR 15-SEP-2000; 2000US-00663870.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX

DR WPI; 2001-476164/51.

DR P-PSDB; AAM24252.

XX

PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.

XX

PS Claim 1; Page 664; 1275pp; English.

XX

CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention

XX

SQ Sequence 580 BP; 146 A; 154 C; 116 G; 164 T; 0 U; 0 Other;

Query Match 11.4%; Score 229.2; DB 4; Length 580;

Best Local Similarity 84.3%; Pred. No. 2.2e-51;

Matches 258; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 1407 ATGTCACCTCGGAGAGGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGTACACTGCTGG 1466

Db	275	AGGTTACTCAGAGAGGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGG	334
Qy	1467	TCCTTATTTCTTTGCCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTCATCAT	1526
Db	335	TCCATATTTCTTTGCCAAGATCCTCGGCGAGCTTCCGGAGCACTGTGCCTACATCATCAT	394
Qy	1527	CTACGCGATGCCCATCTACTGGCTGACAAACCTGCGGCCCGTGCCTGAGCTCTTCCTTCT	1586
Db	395	CTACGGGATGCCACCTACTGGCTGGCCAACTGAGGCCAGGCCTCCAGCCCTTCCTGCT	454
Qy	1587	ACACTTCCTGCTCGTGTGGTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCTGCCTC	1646
Db	455	GCACTTCCTGCTGGAGTGGCTGGCGGTCTTCTGTTGCAAGATTATGGTCCTGGCCGCCGC	514
Qy	1647	TGCCATGCTGCCACCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAACCTCCTT	1706
Db	515	GGGCTGCTCCCCACCTTACACATGGCCTCCTTCTTTCAGCAATGCCCTCTACAACCTGCTT	574
Qy	1707	CTACCT	1712
Db	575	CTACCT	580

ABK51681

ID ABK51681 standard; DNA; 1920 BP.

XX

AC ABK51681;

XX

DT 30-JUL-2002 (first entry)

XX

DE DNA encoding human ABCG5 protein.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW chromosome 2p21; ds.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT CDS 1. .1920

FT /*tag= a

```
FT          /product= "Human ABCG5 protein"
```

```
FT      /transl except= (pos: 4. .9, aa: GDLSSLTPGGSMGL)
```

FT /note= "This sequence contains 13 exons"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.
 XX
 PI Patel ,SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 DR P-PSDB; AAU98984.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Claim 38; Page 36-37; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the human ABCG5 gene located on chromosome 2p21.
 CC This sequence encodes the human ABCG5 protein of the invention
 XX
 SQ Sequence 1920 BP; 440 A; 503 C; 486 G; 491 T; 0 U; 0 Other;

Query Match 9.9%; Score 199.2; DB 6; Length 1920;
 Best Local Similarity 54.0%; Pred. No. 5e-43;
 Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

Qy 234 TAGCAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGG 293
 | || | || | | | | | | | | | | | | | | | | | | |
 Db 141 TTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGG 200
 Qy 294 ACAGATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGAT 353
 |||| | | || | || | || | || | || | || | || | ||
 Db 201 GCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCAT 260
 Qy 354 CACAGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACC 413
 | || | | | | | | | | | | | | | | | | | | | |
 Db 261 GTCCGGGAGGCTGGGGCGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGC 320
 Qy 414 CAGTACGCCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCT 473
 | | || | | | | || | | | | | | | | | | | | |
 Db 321 GCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCTGCT 380

Qy 474 GCCCAACCTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCAGATGCGCCTGCCAGGAC 533
 | | | | | | | | | | | | | | | | | | | | | |
 Db 381 GAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGG 440
 Qy 534 CTTCTCCCAGGCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCG 593
 | | | | | | | | | | | | | | | | | | | | | |
 Db 441 CAATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAG 497
 Qy 594 GCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCG 653
 | | | | | | | | | | | | | | | | | | | | | |
 Db 498 CCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCG 557
 Qy 654 CCGACGAGTGAGCATTTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGA 713
 | | | | | | | | | | | | | | | | | | | | | |
 Db 558 GCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGA 617
 Qy 714 ACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCT 773
 | | | | | | | | | | | | | | | | | | | | | |
 Db 618 GCCAACCACAGGCCCTGGACTGCATGACTGCTAATCAGATTGTGCTCCTCCTGGTGGAAC 677
 Qy 774 GGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAG 833
 | | | | | | | | | | | | | | | | | | | | | |
 Db 678 GGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCCAGCCCCGTTCTGAGCTTTTTCA 737
 Qy 834 GCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCA 893
 | | | | | | | | | | | | | | | | | | | | | |
 Db 738 GCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGC 797
 Qy 894 GCAAATGGTGCACTTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGC 953
 | | | | | | | | | | | | | | | | | | | | | |
 Db 798 GGAAATGCTTGATTTCTTCAATGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTT 857
 Qy 954 GGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCAC 1013
 | | | | | | | | | | | | | | | | | | | | | |
 Db 858 TGACTTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAAC 917
 Qy 1014 CGTGGAGAAGGCACAG 1029
 | | | | | | | | | | | | | | | | | | | | | |
 Db 918 CTCCAAGAGAGTCCAG 933

RESULT 7

AAD22009

ID AAD22009 standard; DNA; 2340 BP.

XX

AC AAD22009;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG).

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21; ds.

XX

OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT CDS 107. .2062
 FT /*tag= a
 FT /product= "Human SSG protein"
 XX
 PN WO200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001WO-US012758.
 XX
 PR 18-APR-2000; 2000US-0198465P.
 PR 15-MAY-2000; 2000US-0204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 XX
 DR WPI; 2002-017598/02.
 DR P-PSDB; AAEL3290.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
 PT useful for screening a compound that increases the level of expression or
 PT activity of SSG polypeptide for treating sterol-related disorder.
 XX
 PS Claim 8; Fig 8; 105pp; English.
 XX
 CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
 CC identifying a compound useful in the treatment or prevention of a sterol-
 CC related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations, and
 CC for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG DNA. Human SSG is located on chromosome
 CC 2p21
 XX
 SQ Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

 Query Match 9.9%; Score 199.2; DB 6; Length 2340;
 Best Local Similarity 54.0%; Pred. No. 5.4e-43;
 Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

 Qy 234 TAGCAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGG 293
 | | | | | | | | | | | | | | | | | | | | | |
 Db 283 TTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGG 342

 Qy 294 ACAGATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGAT 353

```

      |||||  ||  |||| |||| ||||| |||||  ||  ||  ||  ||||  ||
Db      343 GCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCAT 402

Qy      354 CACAGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACC 413
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      403 GTCCGGGAGGCTGGGGCGCGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGC 462

Qy      414 CAGTACGCCTCAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCT 473
      |  |||  |  |  |  |||  |  |  |||  ||||  |||  ||||
Db      463 GCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCT 522

Qy      474 GCCCAACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGAC 533
      |  ||  |||  ||||  |||||  ||  ||  ||  ||  ||  ||  ||  ||
Db      523 GAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGG 582

Qy      534 CTTCTCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCG 593
      |  ||  ||  |  |  ||  |||||  |  ||  ||  |||||  |  |||  |
Db      583 CAATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAG 639

Qy      594 GCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCG 653
      ||  ||  ||  |  |||||  |  ||  ||  |||  |||||  ||
Db      640 CCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCG 699

Qy      654 CCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGA 713
      ||  ||  ||  |||  |  |  ||||  ||  ||  ||  ||  ||  |||  ||
Db      700 GCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGA 759

Qy      714 ACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCT 773
      ||  ||  |||||  |||  ||  ||  ||  ||  |  |  ||  ||  ||  ||
Db      760 GCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCTCCTGGTGGAAC 819

Qy      774 GGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAG 833
      |||  |  |||  |  |  |||  |  ||  ||  |||||  ||  ||||  ||  ||
Db      820 GGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTCA 879

Qy      834 GCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCA 893
      |||  |||||  |  |  |||  |  ||  |  ||  ||  ||  ||  ||  ||
Db      880 GCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGC 939

Qy      894 GCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGC 953
      |  ||||  |  |  ||||  |  ||  ||  |||||  ||  |||||  ||
Db      940 GGAAATGCTTGATTTCTTCAATGACTGCGGTTACCTTGTCTGAACATTCAAACCCTTT 999

Qy      954 GGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCAC 1013
      |||||  |||  ||||  |||  |  ||  |  |  ||||  |||||  ||  ||
Db      1000 TGACTTCTATATGGACCTGACGTGAGTGATACCCAAAGCAAGGAACGGGAAATAGAAAC 1059

Qy      1014 CGTGGAGAAGGCACAG 1029
      |  |||  |  |||
Db      1060 CTCCAAGAGAGTCCAG 1075

```

RESULT 8

AAD48882

ID AAD48882 standard; DNA; 2340 BP.

XX

AC AAD48882;

XX
 DT 24-MAR-2003 (first entry)
 XX
 DE Human ABCG5 DNA.
 XX
 KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 107..2062
 FT /*tag= a
 FT /product= "hABCG5 protein"
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR P-PSDB; AAE31704.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 11; Page 77; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is human ABCG5 DNA
 XX
 SQ Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

Query Match 9.9%; Score 199.2; DB 7; Length 2340;
 Best Local Similarity 54.0%; Pred. No. 5.4e-43;
 Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

Qy	234	TAGCAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGG	293
Db	283	TTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGG	342
Qy	294	ACAGATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGAT	353
Db	343	GCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCAT	402
Qy	354	CACAGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACC	413
Db	403	GTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGC	462
Qy	414	CAGTACGCCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCT	473
Db	463	GCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCT	522
Qy	474	GCCCAACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGAC	533
Db	523	GAGCAGCCTCACCGTGC GCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGG	582
Qy	534	CTTCTCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCG	593
Db	583	CAATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAG	639
Qy	594	GCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCG	653
Db	640	CCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCG	699
Qy	654	CCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGA	713
Db	700	GCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGA	759
Qy	714	ACCCACTTCTGGCCTCGACAGCTTCACAGCCCAATCTGGTGACAACCTTGTCCCGCCT	773
Db	760	GCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTGCTCCTCCTGGTGGAAC	819
Qy	774	GGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAG	833
Db	820	GGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCCAGCCCCGTTCTGAGCTTTTTCA	879
Qy	834	GCTATTTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCA	893
Db	880	GCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGC	939
Qy	894	GCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCTGCTATAGCAACCCTGC	953
Db	940	GGAAATGCTTGATTTCTTCAATGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTT	999
Qy	954	GGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCAC	1013
Db	1000	TGACTTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAAC	1059
Qy	1014	CGTGGAGAAGGCACAG	1029
Db	1060	CTCCAAGAGAGTCCAG	1075

RESULT 9

ABK51682

ID ABK51682 standard; cDNA; 2516 BP.

XX

AC ABK51682;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 cDNA sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
chromosome 2p21; ss.

XX

OS Homo sapiens.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 37-38; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic

CC acid sequence represents the cDNA sequence of human ABCG5 gene located on
CC chromosome 2p21
XX
SQ Sequence 2516 BP; 601 A; 631 C; 636 G; 648 T; 0 U; 0 Other;

Query Match 9.9%; Score 199.2; DB 6; Length 2516;
Best Local Similarity 54.0%; Pred. No. 5.6e-43;
Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

```
Qy      234 TAGCAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGG 293
      | | | | | | | | | | | | | | | | | | | | | |
Db      317 TTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGG 376

Qy      294 ACAGATGTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGAT 353
      | | | | | | | | | | | | | | | | | | | | | |
Db      377 GCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCAT 436

Qy      354 CACAGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACC 413
      | | | | | | | | | | | | | | | | | | | | | |
Db      437 GTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGCCGGGC 496

Qy      414 CAGTACGCCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCT 473
      | | | | | | | | | | | | | | | | | | | | | |
Db      497 GCTGCGCCGGGAGCAGTTCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCT 556

Qy      474 GCCCAACCTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCAGATGCGCCTGCCCAGGAC 533
      | | | | | | | | | | | | | | | | | | | | | |
Db      557 GAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACGCGCTGCTGGCCATCCGCCGCGG 616

Qy      534 CTTCTCCCAGGCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCG 593
      | | | | | | | | | | | | | | | | | | | | | |
Db      617 CAATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAG 673

Qy      594 GCAGTGCCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCG 653
      | | | | | | | | | | | | | | | | | | | | | |
Db      674 CCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCG 733

Qy      654 CCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGA 713
      | | | | | | | | | | | | | | | | | | | | | |
Db      734 GCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGA 793

Qy      714 ACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTGTCCCGCCT 773
      | | | | | | | | | | | | | | | | | | | | | |
Db      794 GCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAC 853

Qy      774 GGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAG 833
      | | | | | | | | | | | | | | | | | | | | | |
Db      854 GGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTCA 913

Qy      834 GCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCA 893
      | | | | | | | | | | | | | | | | | | | | | |
Db      914 GCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGC 973

Qy      894 GCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGC 953
      | | | | | | | | | | | | | | | | | | | | | |
Db      974 GGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCCTGAACATTCAAACCCTTT 1033
```

Qy 954 GGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCAC 1013
 ||||| |||| |||| | || | |||| ||||| | | ||
 Db 1034 TGACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAAC 1093

Qy 1014 CGTGGAGAAGGCACAG 1029
 | || | |||
 Db 1094 CTCCAAGAGAGTCCAG 1109

RESULT 10

ABK51686

ID ABK51686 standard; cDNA; 2035 BP.

XX

AC ABK51686;

XX

DT 07-AUG-2003 (revised)

DT 30-JUL-2002 (first entry)

XX

DE cDNA encoding rat ABCG5 protein.

XX

KW Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; ss;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

XX

OS Rattus sp.

XX

FH Key Location/Qualifiers

FT CDS 8. .1965

FT /*tag= a

FT /product= "Rat ABCG5 protein"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR P-PSDB; AAU96986.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 45-46; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette

CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a

CC predisposition for developing sitosterolemia, arteriosclerosis or heart

CC disease. The molecules of the invention are also useful for identifying a

CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the rat ABCG5 protein of the invention. (Updated on
 CC 07-AUG-2003 to correct OS field.)

XX

SQ Sequence 2035 BP; 481 A; 533 C; 537 G; 484 T; 0 U; 0 Other;

Query Match 9.7%; Score 195; DB 6; Length 2035;
 Best Local Similarity 53.9%; Pred. No. 7e-42;
 Matches 424; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

Qy	261	CATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTC	320
Db	214	CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGACCATGTGCATCTTAGGTAGCTC	273
Qy	321	AGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGAT	380
Db	274	AGGCTCAGGGAAAACCACGCTGCTGGACGCCATCTCTGGGAGGCTGCGGCGCACAGGGAC	333
Qy	381	GAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGTG	440
Db	334	CTTGAAGGGGAAGTGTGTTGTGAACGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTG	393
Qy	441	CGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCCT	500
Db	394	CGTCTCCTACCTCCTGCGAGCGATGTCTTTCTGAGCAGCCTCACGGTGCGGGAGACGCT	453
Qy	501	GGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCAGGCCCAGCGTGACAAACG	560
Db	454	GAGATACACGGC---GATGCTGGCTCTCCGCAGCAGCTCCGCGGACTTCTACGACAAGAA	510
Qy	561	GGTGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAA	620
Db	511	GGTAGAGGCAGTCTTGACAGAGCTGAGTCTGAGCCACGTGGCAGACCAATGATCGGCAA	570
Qy	621	CACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCT	680
Db	571	CTATAATTTTGGGGGATTTCAGTGCGGAGCGGCGCCGAGTGTCCATCGCAGCCCAACT	630
Qy	681	CCTGTGAACCCAGGAATCCTCATTCTGGATGAACCACTTCTGGCCTCGACAGCTTCAC	740
Db	631	CCTTCAGGACCCCAAGGTCATGATGCTTGACGAGCCAACCACAGGACTGGACTGCATGAC	690
Qy	741	AGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCAT	800

Db	691	TGCAATCATATCGTCCTCCTCTTGGTCGAGCTGGCTCGCAGGAACCGCATTGTAATTGT	750
Qy	801	CTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGAC	860
Db	751	CACCATCCACCAGCCTCGCTCTGAGCTCTTCCACCACTTCGACAAAATTGCCATTCTGAC	810
Qy	861	ATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCAT	920
Db	811	TTACGGAGAGTTGGTGTTCTGTGGCACGCCAGAGGAGATGCTCGGCTTCTTCAATAACTG	870
Qy	921	TGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCAT	980
Db	871	TGGTTACCCCTGTCCTGAACATTCCAATCCCTTTGATTTCTACATGGACTTGACATCGGT	930
Qy	981	CGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAGAAGGCACAGTCTCTTGACAGC	1040
Db	931	GGACACCCAAAGCAGAGAGCGAGAGATAGAGACGTACAAGCGAGTCCAGATGCTGGAATC	990
Qy	1041	CCTGTTT	1047
Db	991	TGCCTTC	997

Db 387 CTTCTCCTACGTCCTGCAGAGCGACGTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTT 446

Qy 501 GGCTTTCATTGCCCAGATGCGCCTGCCAGGACCTTCTCCCAGGCCAGCGTGACAAACG 560
 | | | | | | | | | | | | | | | | | | | | | |
 Db 447 GCGATACACAGC---GATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTACAACAAGAA 503
 Qy 561 GGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAA 620
 | | | | | | | | | | | | | | | | | | | | | |
 Db 504 GGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAATGATTGGCAG 563
 Qy 621 CACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCT 680
 | | | | | | | | | | | | | | | | | | | | | |
 Db 564 CTATAATTTTGGGGGAATTTCCAGTGCGGAGCGGCGCCGAGTTTCCATCGCAGCCCAACT 623
 Qy 681 CCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCAC 740
 | | | | | | | | | | | | | | | | | | | | | |
 Db 624 CCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGAC 683
 Qy 741 AGCCCAACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCAT 800
 | | | | | | | | | | | | | | | | | | | | | |
 Db 684 TGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGT 743
 Qy 801 CTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCCTTCTGATGAC 860
 | | | | | | | | | | | | | | | | | | | | | |
 Db 744 CACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGAC 803
 Qy 861 ATCTGGCACCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCAT 920
 | | | | | | | | | | | | | | | | | | | | | |
 Db 804 TTACGGAGAGTTGGTGTCTGTGGCACCCCAAGGAGATGCTTGGCTTCTTCAATAACTG 863
 Qy 921 TGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCAT 980
 | | | | | | | | | | | | | | | | | | | | | |
 Db 864 TGGTTACCCCTGTCCTGAACATTCCAATCCCTTTGATTTTACATGGACTTGACATCAGT 923
 Qy 981 CGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAGAAGGCACAGTCTCTTGACG 1040
 | | | | | | | | | | | | | | | | | | | | | |
 Db 924 GGACACCCAAAGCAGAGAGCGGGAATAGAAACGTACAAGCGAGTACAGATGCTGGAATG 983
 Qy 1041 CCTGTTTCCTAGAA 1053
 | | | | | |
 Db 984 TGCCTTCAAGGAA 996

RESULT 12

AAD48880

ID AAD48880 standard; DNA; 1959 BP.

XX

AC AAD48880;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG5 DNA.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5; gene; ds.

XX

OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1. .1591
 FT /*tag= a
 FT /product= "mABCG5 protein"
 XX

PN WO200281691-A2.

PD 17-OCT-2002.

PF 20-NOV-2001; 2001WO-US043823.

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

PA (TULA-) TULARIK INC.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Hobbs HH, Shan B, Barnes R, Tian H;

DR WPI; 2003-058548/05.

DR P-PSDB; AAE31702.

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-related disorders e.g. sitosterolemia, hypercholesterolemia, hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or nutritional deficiencies.

PS Claim 11; Page 73; 94pp; English.

CC The invention relates to ATP-binding cassette (ABC) family cholesterol transporter, ABCG8 polypeptides and polynucleotides. The invention also provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known as sitosterolaemia susceptibility gene (SSG). Sequences of the invention are useful for treating or preventing sterol-related disorders such as sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis and nutritional deficiencies. They are also useful in gene therapy. The present sequence is mouse ABCG5 DNA

SQ Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 U; 0 Other;

Query Match 9.2%; Score 186.6; DB 7; Length 1959;
 Best Local Similarity 53.1%; Pred. No. 1.3e-39;
 Matches 421; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

Qy	261	CATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTC	320
Db	207	CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTC	266
Qy	321	AGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGAT	380
Db	267	AGGCTCAGGGAAGACCACGTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGAC	326
Qy	381	GAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGTG	440
Db	327	CCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAAGACTG	386

Qy 441 CGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCCT 500
 | | | | | | | | | | | | | | | | | | | | | |
 Db 387 CTTCTCCTACGTCTGCGAGCGACGTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTT 446

 Qy 501 GGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCAGCGTGACAAACG 560
 | | | | | | | | | | | | | | | | | | | | | |
 Db 447 GCGATACACAGC---GATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTACAACAAGAA 503

 Qy 561 GGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAA 620
 | | | | | | | | | | | | | | | | | | | | | |
 Db 504 GGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAATGATTGGCAG 563

 Qy 621 CACGTATGTACGTGGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCT 680
 | | | | | | | | | | | | | | | | | | | | | |
 Db 564 CTATAATTTTGGGGGAATTTCCAGTGCGGAGCGGCGCCGAGTTTCCATCGCAGCCCAACT 623

 Qy 681 CCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCAC 740
 | | | | | | | | | | | | | | | | | | | | | |
 Db 624 CCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGAC 683

 Qy 741 AGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCAT 800
 | | | | | | | | | | | | | | | | | | | | | |
 Db 684 TGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGT 743

 Qy 801 CTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGAC 860
 | | | | | | | | | | | | | | | | | | | | | |
 Db 744 CACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGAC 803

 Qy 861 ATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCAT 920
 | | | | | | | | | | | | | | | | | | | | | |
 Db 804 TTACGGAGAGTTGGTGTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTTCAATAACTG 863

 Qy 921 TGGCCACCCTTGTCCCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCAT 980
 | | | | | | | | | | | | | | | | | | | | | |
 Db 864 TGGTTACCCCTGTCTGAACATTCCAATCCCTTTGATTTTTACATGGACTTGACATCAGT 923

 Qy 981 CGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAGAAGGCACAGTCTCTTGCAGC 1040
 | | | | | | | | | | | | | | | | | | | | | |
 Db 924 GGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTACAAGCGAGTACAGATGCTGGAATG 983

 Qy 1041 CCTGTTCTTAGAA 1053
 | | | | |
 Db 984 TGCCTTCAAGGAA 996

RESULT 13

AAD22008

ID AAD22008 standard; DNA; 2258 BP.

XX

AC AAD22008;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG).

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 47. .2005
FT /*tag= a
FT /product= "Mouse SSG protein"
XX
PN WO200179272-A2.
XX
PD 25-OCT-2001.
XX
PF 18-APR-2001; 2001WO-US012758.
XX
PR 18-APR-2000; 2000US-0198465P.
PR 15-MAY-2000; 2000US-0204234P.
XX
PA (TULA-) TULARIK INC.
XX
PI Tian H, Schultz J, Shan B;
XX
DR WPI; 2002-017598/02.
DR P-PSDB; AAE13289.
XX
PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
PT useful for screening a compound that increases the level of expression or
PT activity of SSG polypeptide for treating sterol-related disorder.
XX
PS Claim 8; Fig 7; 105pp; English.
XX
CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
CC identifying a compound useful in the treatment or prevention of a sterol-
CC related disorder, including sitosterolaemia, hyperlipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolaemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolaemia and other
CC cardiovascular disorders, for forensics and paternity determinations, and
CC for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG DNA. Mouse SSG is located on chromosome 17
XX
SQ Sequence 2258 BP; 549 A; 579 C; 567 G; 563 T; 0 U; 0 Other;

Query Match 9.2%; Score 186.6; DB 6; Length 2258;
Best Local Similarity 53.1%; Pred. No. 1.4e-39;
Matches 421; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

Qy 261 CATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTC 320

ABK51685

ID ABK51685 standard; cDNA; 2354 BP.

XX

AC ABK51685;

XX

DT 30-JUL-2002 (first entry)

XX

DE Mouse ABCG5 cDNA sequence.

XX

KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW ss.

XX

OS Mus sp.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 45; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of the

CC invention

XX

SQ Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 U; 0 Other;

Query Match 9.2%; Score 186.6; DB 6; Length 2354;

Best Local Similarity 53.1%; Pred. No. 1.4e-39;

Matches 421; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

```
Qy      261 CATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTC 320
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      345 CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTC 404

Qy      321 AGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGAT 380
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      405 AGGCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGAC 464

Qy      381 GAAATCAGGACAAATTTGGATAAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGTG 440
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      465 CCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTG 524

Qy      441 CGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCCT 500
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      525 CTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCAGACGTT 584

Qy      501 GGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCAGCGTGACAAACG 560
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      585 GCGATACACAGC---GATGCTGGCCCTCTGCCGCAGCTCCGCGGACTTCTACAACAAGAA 641

Qy      561 GGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAA 620
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      642 GGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAATGATTGGCAG 701

Qy      621 CACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCT 680
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      702 CTATAATTTTGGGGGAATTTCCAGTGCGGAGCGGCGCCGAGTTTCCATCGCAGCCCAACT 761

Qy      681 CCTGTGGAACCCAGGAATCCTCATTTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCAC 740
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      762 CCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGAC 821

Qy      741 AGCCCACAATCTGGTGACAACCTTGTCCCGCTGGCCAAGGGCAACAGGCTGGTGCTCAT 800
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      822 TGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGT 881

Qy      801 CTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCTCTTGATGAC 860
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      882 CACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCTTGAC 941

Qy      861 ATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCAT 920
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      942 TTACGGAGAGTTGGTGTTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTTCAATAACTG 1001

Qy      921 TGGCCACCCTTGTCCTCGCTATAGCAACCTGCGGACTTCTACGTGGACTTGACCAGCAT 980
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1002 TGGTTACCCCTGTCCTGAACATTCCAATCCCTTTGATTTTTACATGGACTTGACATCAGT 1061

Qy      981 CGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAGAAGGCACAGTCTCTTGCAGC 1040
```

```

      ||| | | ||| ||| |||| | | || | | ||| | | | |
Db      1062 GGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTACAAGCGAGTACAGATGCTGGAATG 1121

Qy      1041 CCTGTTCTCTAGAA 1053
      ||| |||
Db      1122 TGCCTTCAAGGAA 1134

```

RESULT 15

ABK51687

ID ABK51687 standard; cDNA; 1069 BP.

XX

AC ABK51687;

XX

DT 07-AUG-2003 (revised)

DT 30-JUL-2002 (first entry)

XX

DE cDNA encoding hamster ABCG5 protein.

XX

KW Hamster; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW ss.

XX

OS Cricetinae.

XX

FH Key Location/Qualifiers

FT CDS 30..1049

FT /*tag= a

FT /partial

FT /product= "Hamster ABCG5 protein"

FT /note= "This sequence lacks both a start and stop codon"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR P-PSDB; AAU96987.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 47; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette

CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a

CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the hamster ABCG5 protein of the invention.
 CC (Updated on 07-AUG-2003 to correct OS field.)

XX

SQ Sequence 1069 BP; 266 A; 282 C; 273 G; 248 T; 0 U; 0 Other;

Query Match 8.7%; Score 176; DB 6; Length 1069;
 Best Local Similarity 56.5%; Pred. No. 7e-37;
 Matches 348; Conservative 0; Mismatches 265; Indels 3; Gaps 1;

Qy	437	AGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGA	496
Db	118	ACTGCTTCTCCTATGTCTGCGAGCGACGCTCTTCTGAGCAGTCTCACGGTGCGAGAGA	177
Qy	497	CCCTGGCTTTTCATTGCCAGATGCGCCTGCCAGGACCTTCTCCAGGCCAGCGTGACA	556
Db	178	CGCTGCGCTACACGGCGATGCTGGCCCTCCGCGAGTAGCTCTTCGGACTTCTA---TGACA	234
Qy	557	AACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGG	616
Db	235	AGAAGGTAGAGGCAGTCATGGAAGAGCTAAGTCTGAGCCACGTGGCAGACCGAATGATTG	294
Qy	617	GCAACACGTATGTACGTGGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGC	676
Db	295	GCAACTATAATTTTGGGGGAATTTCCAGTGCGGAGCGGCGCCGAGTCTCCATCGCAGCCC	354
Qy	677	AGCTCCTGTGGAACCCAGGAATCCTCATTTCTGGATGAACCCACTTCTGGCCTCGACAGCT	736
Db	355	AACTCATTGAGGACCCCAAGATCATGATGTTTGATGAGCCAACACAGGACTGGACTGCA	414
Qy	737	TCACAGCCCACAATCTGGTGACAACCTTGTCGCCGCTGGCCAAGGGCAACAGGCTGGTGC	796
Db	415	TGACTGCAAATCAAATTGTCATCCTCCTGGCAGAGCTGGCTCGCAGGGACCGCATTGTGA	474
Qy	797	TCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGA	856
Db	475	TCGTCACCATCCACCAGCCTCGCTCTGAGCTCTTTCAACACTTCGACAAAATTGCCATCC	534
Qy	857	TGACATCTGGCACCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACAT	916
Db	535	TGACTTACGGAGAGATGGTGTCTGTGGCACGCCGAGGAAATGCTCGACTTCTTCAATA	594

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:48:03 ; Search time 97.675 Seconds
(without alignments)
11471.161 Million cell updates/sec

Title: US-09-989-981A-3
Perfect score: 2019
Sequence: 1 atggctgagaaaaccaaaga.....agtcaattcaagactgggtga 2019

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		%					
Result		Query					Description
No.	Score	Match	Length	DB	ID		
1	132.4	6.6	2418	4	US-09-245-808-2		Sequence 2, Appli
2	63.4	3.1	7218	1	US-08-232-463-14		Sequence 14, Appl
3	62.2	3.1	4159	4	US-09-614-912-139		Sequence 139, App
4	59.4	2.9	3376	4	US-09-620-312D-918		Sequence 918, App
5	55	2.7	1977	4	US-09-614-912-143		Sequence 143, App
6	52.8	2.6	4403765	3	US-09-103-840A-2		Sequence 2, Appli
7	52.8	2.6	4411529	3	US-09-103-840A-1		Sequence 1, Appli
8	51	2.5	2031	4	US-09-614-912-137		Sequence 137, App
c 9	49.8	2.5	630	4	US-09-489-039A-932		Sequence 932, App
10	49.8	2.5	960	4	US-09-489-039A-945		Sequence 945, App
c 11	49.6	2.5	627	4	US-09-252-991A-12021		Sequence 12021, A

	12	49.6	2.5	732	4	US-09-252-991A-11963	Sequence 11963, A
	13	49.6	2.5	2328	4	US-09-252-991A-11890	Sequence 11890, A
c	14	49.4	2.4	705	4	US-09-252-991A-12050	Sequence 12050, A
c	15	48.6	2.4	28804	2	US-08-592-874-1	Sequence 1, Appli
c	16	48.6	2.4	28804	3	US-09-096-942-2	Sequence 2, Appli
c	17	48.6	2.4	28804	3	US-09-096-867-2	Sequence 2, Appli
	18	48.4	2.4	876	4	US-09-489-039A-1001	Sequence 1001, Ap
	19	46.8	2.3	996	4	US-09-252-991A-2920	Sequence 2920, Ap
c	20	46.8	2.3	1284	4	US-09-252-991A-3041	Sequence 3041, Ap
	21	46.8	2.3	1476	4	US-09-252-991A-2825	Sequence 2825, Ap
	22	45.8	2.3	723	4	US-09-252-991A-11541	Sequence 11541, A
c	23	45.8	2.3	1155	4	US-09-252-991A-11845	Sequence 11845, A
	24	45.8	2.3	2367	4	US-09-252-991A-11600	Sequence 11600, A
c	25	44.6	2.2	435	4	US-09-252-991A-9969	Sequence 9969, Ap
	26	44.6	2.2	900	4	US-09-252-991A-10183	Sequence 10183, A
c	27	44.6	2.2	1332	4	US-09-252-991A-9889	Sequence 9889, Ap
	28	44.4	2.2	2752	1	US-08-430-925A-3	Sequence 3, Appli
	29	44	2.2	3343	4	US-09-976-594-502	Sequence 502, App
	30	43.6	2.2	1050	4	US-09-489-039A-3086	Sequence 3086, Ap
c	31	43	2.1	276	4	US-09-252-991A-4049	Sequence 4049, Ap
	32	43	2.1	417	4	US-09-252-991A-3926	Sequence 3926, Ap
	33	43	2.1	606	4	US-09-252-991A-3898	Sequence 3898, Ap
c	34	43	2.1	765	4	US-09-252-991A-3980	Sequence 3980, Ap
	35	43	2.1	1047	4	US-08-540-650B-6	Sequence 6, Appli
	36	43	2.1	1053	4	US-09-016-434-1423	Sequence 1423, Ap
	37	43	2.1	1882	4	US-08-540-650B-11	Sequence 11, Appl
c	38	43	2.1	2223	4	US-09-252-991A-4015	Sequence 4015, Ap
	39	43	2.1	3083	4	US-08-693-308-1	Sequence 1, Appli
	40	42.8	2.1	732	4	US-09-489-039A-6450	Sequence 6450, Ap
c	41	42.4	2.1	411	4	US-09-252-991A-5107	Sequence 5107, Ap
	42	42.4	2.1	1875	4	US-09-252-991A-5054	Sequence 5054, Ap
	43	42.4	2.1	1962	4	US-09-252-991A-5020	Sequence 5020, Ap
c	44	42.4	2.1	2295	4	US-09-252-991A-5162	Sequence 5162, Ap
	45	42.4	2.1	35081	2	US-08-752-760A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-245-808-2

; Sequence 2, Application US/09245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; TITLE OF INVENTION: encodes it

; FILE REFERENCE: Ross UMb conversion

; CURRENT APPLICATION NUMBER: US/09/245,808

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-2

Query Match 6.6%; Score 132.4; DB 4; Length 2418;
Best Local Similarity 51.9%; Pred. No. 1e-26;
Matches 348; Conservative 0; Mismatches 316; Indels 6; Gaps 2;

```
Qy      304 GCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGA 363
          ||||| | | | | |||| | | | | | | | | | | | | | |
Db      467 GCCATCCTGGGACCCACAGGTGGAGGCAAATCTTCGTTATTAGATG---TCTTAGCTGCA 523

Qy      364 GGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCT 423
          | | | | | | | | |||| | | | | | | | | | | | |
Db      524 AGGAAAGATCCAAGTGGATTATCTGGAGATGTTCTGATAAATGGAGCACCGCGACCTGCC 583

Qy      424 CAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTG 483
          | | | | | | | | | |||| | | | | | | | | | |
Db      584 AATTTCAAATGTAATTCAGGT---TACGTGGTACAAGATGATGTTGTGATGGGCACTCTG 640

Qy      484 ACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAG 543
          || || |||| | | | | || | | | | | | | | | |
Db      641 ACGGTGAGAGAAAACCTTACAGTTCTCAGCAGCTCTTCGGCTTGCAACAACCTATGACGAAT 700

Qy      544 GCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCC 603
          | | | | | | | | | | | | | | | | | | | |
Db      701 CATGAAAAAACGAACGGATTAACAGGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGA 760

Qy      604 AACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTG 663
          || || | | | | | | | | | | | | | | | | | |
Db      761 GACTCCAAGGTTGGAACCTCAGTTTATCCGTGGTGTGTCTGGAGGAGAAAGAAAAGGACT 820

Qy      664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
          || || | | | | | | | | | | | | | | | | | |
Db      821 AGTATAGGAATGGAGCTTATCACTGATCCTTCCATCTTGTTCTTGGATGAGCCTACAAC 880

Qy      724 GGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGC 783
          || | | | | | | | | | | | | | | | | | |
Db      881 GGCTTAGACTCAAGCACAGCAAATGCTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAG 940

Qy      784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
          | | | | | | | | | | | | | | | | | |
Db      941 GGACGAACAATCATCTTCTCCATTCATCAGCCTCGATATTCCATCTTCAAGTTGTTTGAT 1000

Qy      844 CTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTG 903
          || | | | | | | | | | | | | | | | | |
Db      1001 AGCCTCACCTTATTGGCCTCAGGAAGACTTATGTTCACGGGCTGCTCAGGAGGCCTTG 1060

Qy      904 CAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTAC 963
          |||| | | | | | | | | | | | | | | | |
Db      1061 GGATACTTTGAATCAGCTGGTTATCACTGTGAGGCCTATAATAACCCTGCAGACTTCTTC 1120

Qy      964 GTGGACTTGA 973
          |||| | |
Db      1121 TTGGACATCA 1130
```

US-08-232-463-14

US-08-232-463-14

Qy 1567 GTGCCTGAGCTCTTCTTCTACACTTCCTGCTCGTGGGTTGGTGGTCTTCTGCTGCAGG 1626
 | ||| | | ::::: : ::::: ::: : : :: : ::::: :: :
Db 1056 GAGCTTGCGATYYY 1115

; PRIOR APPLICATION NUMBER: 60/172,959
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/172,946
 ; PRIOR FILING DATE: 1999-12-21
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 139
 ; LENGTH: 4159
 ; TYPE: DNA
 ; ORGANISM: Oryza sativa
 US-09-614-912-139

Query Match 3.1%; Score 62.2; DB 4; Length 4159;
 Best Local Similarity 49.1%; Pred. No. 4.5e-07;
 Matches 194; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

Qy	583	CTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCC	642
Db	396	CTGGGATTGGATATATGCGCGGACACGATCGTCGGCGACCAGATGCAGAGGGGGATCTCC	455
Qy	643	GGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTC	702
Db	456	GGTGGTCAGAAGAAACGCGTCACCACCGGTGAGATGATTGTCTGGTCCAACAAAGGTTCTA	515
Qy	703	ATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACC	762
Db	516	TTCATGGATGAGATATCAACTGGATTGGACAGCTCCACCACATTCCAGATTGTCAAATGC	575
Qy	763	TTGTCCCG---CCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGC	819
Db	576	CTTCAGCAAATCGTGCACTTGGGGCGAGGCAACCATCCTCATGTCACTCCTACAACCAGCC	635
Qy	820	TCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTAC	879
Db	636	CCTGAGACTTTTGAGCTATTCGATGACATTATCCTACTGTCAGAAGGCCAGATTGTTTAT	695
Qy	880	CTGGGGGGCGGCGCAGCAAATGGTGCACTTTCACATCCATTGGCCACCCTTGTCCTCGC	939
Db	696	CAGGGACCCCGCAATACGTCCTTGAGTTCTTTGAGTCATGCGGATTCCGCTGCCCAGAG	755
Qy	940	TATAGCAACCCTGCGGACTTCTACGTGGACTTGAC	974
Db	756	CGTAAGGTACTGCAGACTTCTTCAGGAGGTGAC	790

RESULT 4

US-09-620-312D-918
 ; Sequence 918, Application US/09620312D
 ; Patent No. 6569662
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.


```

; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 918
; LENGTH: 3376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2808)
US-09-620-312D-918

```

```

Query Match          2.9%; Score 59.4; DB 4; Length 3376;
Best Local Similarity 48.4%; Pred. No. 2.4e-06;
Matches 279; Conservative 0; Mismatches 271; Indels 27; Gaps 3;

```

```

Qy      280 AAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCA 339
      ||| | | | | || || | | ||||| || ||||| || | | ||
Db      88 AAATTCTGCCGCCGGGAGCTGATTGGCATCATGGGCCCTCAGGGGCTGGCAAGTCTACA 147

Qy      340 CTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGG 399
      | | || | | |||| | | ||| || | | || || || | |
Db      148 TTCATGAACATCTTGGCAGGATACAGGGAGTCTGGAATGAAG-----GGGCAGATCCTG 201

Qy      400 ATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAG 459
      | |||| | | | || | | | | || | | | | | | | |
Db      202 GTTAATGGAAGGCCACGGGAGCTGAGGACCTTCCGCAAGATGTCCTGCTACATCATGCAA 261

Qy      460 CATGACCAACTGCTGCCCCAACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATG 519
      |||| | ||||| |||| || || || || || || || || ||
Db      262 GATGACATGCTGCTGCCGCACCTCACGGTGTTGGAAGCCATGATGGTCTCTGCTAACCTG 321

Qy      520 CGCCTGCCCAGGACCTTCTCCCAGGCCAGCGTGACAAACGGGTGGAAGACGTAATCGCC 579
      ||| || |||| | | | | |||| || | | | |
Db      322 AAGCTGAGTGAGA-----AGCAGGAGGTGAAGAAGGAGCTGGTGACAGAGATCCTGACG 375

Qy      580 GAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTG 639
      | ||| | || | |||| || |||| || | | | | | |
Db      376 GCACTGGGCCTGATGTCGTGCTCCACACGAGGACAGCC-----CTGCTC 420

```

Qy 640 TCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATC 699
 || || || || | || || || | || || | || || ||
 Db 421 TCTGGCGGGCAGAGGAAGCGTCTGGCCATCGCCCTGGAGCTGGTCAACAACCCGCCTGTC 480
 Qy 700 CTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACA 759
 | | | |||| |||| || || || || | | | |||| |
 Db 481 ATGTTCTTTGATGAGCCACCACTGGTCTGGATAGCGCCTCTTGTTCCTCAAGTGGTGTCC 540
 Qy 760 ACCTTGTCCTCGCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGC 819
 | || |||| || || | | | || || || || || || ||
 Db 541 CTCATGAAGTCCCTGGCACAGGGGGGCGGTACCATCATCTGCACCATCCACCAGCCCACT 600
 Qy 820 TCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGA 856
 | | |||| | |||| | | | |
 Db 601 GCCAAGCTCTTTGAGATGTTTGACAAGTGCATCTTCA 637

RESULT 5

US-09-614-912-143

; Sequence 143, Application US/09614912
 ; Patent No. 6677502
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Steve
 ; APPLICANT: Rafalski, Antoni
 ; APPLICANT: Orozco, Buddy
 ; APPLICANT: Miao, Gou-Hau
 ; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Lee, Jian Ming
 ; APPLICANT: Sakai, Hajime
 ; APPLICANT: Weng, Zude
 ; APPLICANT: Caimi, Perry G
 ; APPLICANT: Anderson, Shawn
 ; TITLE OF INVENTION: Plant Metabolism Genes
 ; FILE REFERENCE: BB1378 US NA
 ; CURRENT APPLICATION NUMBER: US/09/614,912
 ; CURRENT FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,401
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,412
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/146,650
 ; PRIOR FILING DATE: 1999-07-30
 ; PRIOR APPLICATION NUMBER: 60/170,906
 ; PRIOR FILING DATE: 1999-12-15
 ; PRIOR APPLICATION NUMBER: 60/172,959
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/172,946
 ; PRIOR FILING DATE: 1999-12-21
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 143
 ; LENGTH: 1977
 ; TYPE: DNA
 ; ORGANISM: Triticum aestivum
 US-09-614-912-143

Query Match 2.7%; Score 55; DB 4; Length 1977;
 Best Local Similarity 51.4%; Pred. No. 3.1e-05;
 Matches 127; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

Qy      613 GTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGG 672
        || || | | | | | | | | | | | | | | | | | | | |
Db      248 GTTGGGCTCCCTGGAGTGAATGGTCTATCAACTGAGCAACGCAAGAGGCTTACAATTGCC 307

Qy      673 GTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGAC 732
        ||| ||| | | | | | | | | | | | | | | | | | | |
Db      308 GTGGAGCTTGTGCTAACCCGTCGATCATTTTTATGGATGAGCCAACATCTGGTCTTGAT 367

Qy      733 AGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTG 792
        ||| | | | | | | | | | | | | | | | | | | |
Db      368 GCTCGTGCAGCTGCAATTGTGATGAGGACTGTTAGGAACACTGTTAACACTGGCAGGACC 427

Qy      793 GTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTT 852
        || | | | | | | | | | | | | | | | | | | | |
Db      428 GTTGTTCGACCATCCACCAGCCAAGTATTGACATATTTGAAGCATTGATGAGCTTTTC 487

Qy      853 CTGATGA 859
        |||||
Db      488 TTGATGA 494
  
```

RESULT 6

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 2.6%; Score 52.8; DB 3; Length 4403765;
 Best Local Similarity 49.4%; Pred. No. 0.0067;
 Matches 178; Conservative 0; Mismatches 167; Indels 15; Gaps 1;

```

Qy      451 GTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCCTGGCTTTTCATT 510
        |||| ||| | ||| ||| | ||||| | | | | ||| |
  
```

Db 1965645 GTGCCACAGGACGACGTGGTGCACGGTCAGCTGACCGTGAAACACGCGCTGATGTATGCC
1965704

Qy 511 GCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCCAGCGTGACAAACGGGTGGAAGAC 570
| | | | | | | | | | | | | | | | | | | | | |

Db 1965705 GCCGAACCTACGGCTGCCGCCGACACCACCAAAGATGACCGCACCCAGGTAGTTGCCCGG
1965764

Qy 571 GTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTA 630
| | | | | | | | | | | | | | | | | | | | | |

Db 1965765 GTGCTCGAAGAACTCGAGATGTCCAAGCACATCGACACCAGGGTCGACAA-----
1965814

Qy 631 CGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAAC 690
| | | | | | | | | | | | | | | | | | | | | |

Db 1965815 -----GCTGTCGGGTGGTCAACGCAAGCGGGCGTCGGTGGCGCTTGAGCTGTTGACCGG
1965869

Qy 691 CCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAAT 750
| | | | | | | | | | | | | | | | | | | | | |

Db 1965870 CCGTCACTGCTGATCCTCGACGAGCCGACATCCGGCCTAGATCCTGCGCTGGACCGGCAG
1965929

Qy 751 CTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCAC 810
| | | | | | | | | | | | | | | | | | | | | |

Db 1965930 GTCATGACCATGCTGCGGCAGTTGGCCGACGCCGGTCGGGTGGTGCTCGTGGTTACCCAC
1965989

RESULT 7

US-09-103-840A-1

; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 2.6%; Score 52.8; DB 3; Length 4411529;
Best Local Similarity 49.4%; Pred. No. 0.0067;
Matches 178; Conservative 0; Mismatches 167; Indels 15; Gaps 1;

Qy 451 GTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCCTGGCTTTCATT 510

```

      |||| | | | | | | | | | | | | | | | |
Db    1974816 GTGCCACAGGACGACGTGGTGCACGGTCAGCTGACCGTGAAACACGCGCTGATGTATGCC
1974875

Qy      511 GCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCCAGCGTGACAAACGGGTGGAAGAC 570
      ||| | | | | | | | | | | | | | | |
Db    1974876 GCCGAACCTACGGCTGCCGCCGGACACCACCAAAGATGACCGCACCCAGGTAGTTGCCCGG
1974935

Qy      571 GTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTA 630
      || ||| || || | || | | | | | | | |
Db    1974936 GTGCTCGAAGAACTCGAGATGTCCAAGCACATCGACACCAGGGTCGACAA-----
1974985

Qy      631 CGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAAC 690
      | |||| || ||| | ||| || | | | | | | ||
Db    1974986 -----GCTGTCGGGTGGTCAACGCAAGCGGGCGTCGGTGGCGCTTGAAGCTGTTGACCGG
1975040

Qy      691 CCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAAT 750
      || | | | | | | | | | | | | | | | | |
Db    1975041 CCGTCACTGCTGATCCTCGACGAGCCGACATCCGGCCTAGATCCTGCGCTGGACCGGCAG
1975100

Qy      751 CTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCAC 810
      | |||| | || | |||| | | | | | | | | |
Db    1975101 GTCATGACCATGCTGCGGCAGTTGGCCGACGCCGGTCGGGTGGTGCTCGTGGTTACCCAC
1975160

```

RESULT 8

US-09-614-912-137

```

; Sequence 137, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15

```

```
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 137
; LENGTH: 2031
; TYPE: DNA
; ORGANISM: Zea mays
US-09-614-912-137
```

RESULT 9

; LENGTH: 630
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-932

Query Match 2.5%; Score 49.8; DB 4; Length 630;
Best Local Similarity 46.4%; Pred. No. 0.00048;
Matches 162; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

```
Qy      558 ACGGGTGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGG 617
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      476 AAGGATCGCCGACCGGATCGACGAGCTGATGGCGCTGCTGGGGCTGGAGGCGACGCTGCG 417

Qy      618 CAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCA 677
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      416 CGACCGTTACCCGCATCAGCTCTCCGGCGGCCAGCAGCAGCGGGTGGGGGTGGCGCGGGC 357

Qy      678 GCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTT 737
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      356 GCTGGCGGCAGATCCGGAGGTGCTGTTGATGGATGAGCCCTTCGGCGCCCTCGACCCGGT 297

Qy      738 CACAGCCCAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCT 797
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      296 GACCCGCGAGGCGCTGCAGCAGGAGATGCTGCGCATCCACCGTCTGCTGGGACGGACGAT 237

Qy      798 CATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGAT 857
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      236 TGTGCTGGTGACCCATGATATTGACGAAGCGCTGCGTCTGGCGGACCACCTGGTGCTGAT 177

Qy      858 GACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCG 906
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      176 GGACGGGGGCGAGGTGGTCCAGCAGGGGGCGCCGCTGGAGATGCTCCTG 128
```

RESULT 10

US-09-489-039A-945
; Sequence 945, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 945
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-945

Query Match 2.5%; Score 49.8; DB 4; Length 960;
Best Local Similarity 46.4%; Pred. No. 0.0006;

Matches 162; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

```

Qy      558 ACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGG 617
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      336 AAGGATCGCCGACCGGATCGACGAGCTGATGGCGCTGCTGGGGCTGGAGGCGACGCTGCG 395

Qy      618 CAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCA 677
      | | | | | | | | | | | | | | | | | | | | | |
Db      396 CGACCGTTACCCGCATCAGCTCTCCGGCGGCCAGCAGCAGCGGGTGGGGGTGGCGCGGGC 455

Qy      678 GCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTT 737
      | | | | | | | | | | | | | | | | | | | | | |
Db      456 GCTGGCGGCAGATCCGGAGGTGCTGTTGATGGATGAGCCCTTCGGCGCCCTCGACCCGGT 515

Qy      738 CACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCT 797
      | | | | | | | | | | | | | | | | | | | | | |
Db      516 GACCCGCGAGGCGCTGCAGCAGGAGATGCTGCGCATCCACCGTCTGCTGGGACGGACGAT 575

Qy      798 CATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGAT 857
      | | | | | | | | | | | | | | | | | | | | | |
Db      576 TGTGCTGGTGACCCATGATATTGACGAAGCGCTGCGTCTGGCGGACCACCTGGTGCTGAT 635

Qy      858 GACATCTGGCAGCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGAG 906
      | | | | | | | | | | | | | | | | | | | | | |
Db      636 GGACGGGGGCGAGGTGGTCCAGCAGGGGGGCGCCGCTGGAGATGCTCCTG 684

```

RESULT 11

US-09-252-991A-12021/c

; Sequence 12021, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12021

; LENGTH: 627

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12021

Query Match 2.5%; Score 49.6; DB 4; Length 627;

Best Local Similarity 51.3%; Pred. No. 0.00055;

Matches 115; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```

Qy      591 GCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGA 650
      | | | | | | | | | | | | | | | | | | | | | |
Db      583 GCAGCGCTACGGCATGCCGCTGGAGCCTCGCCGGCTGGTCCATGGGCTGTCCATCGGCGA 524

```



```

Qy      651 GCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGA 710
        ||||  || |||  || |  ||  ||  ||  |  ||  ||  ||  ||
Db      523 GCGCCAGCGGGTGGAGATCGTGCCTGCCTGATGCAGGACATCCGCCTGCTGATCCTCGA 464

Qy      711 TGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCG 770
        || || ||||  | ||  |  ||  |  |||  |  |||  |  ||  ||  ||
Db      463 CGAGCCGACTTCGGTGCTGACCCACGCGAGGCCGAGGATCTCTTCGTCACCCTGCGCCG 404

Qy      771 CCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGC 814
        || ||  |  |  |  ||  ||  ||  |  |||  ||
Db      403 TCTTGCGGAAGAGGGCTGCAGTGTCTCTTCATCAGCCACAAGC 360

```

RESULT 12

US-09-252-991A-11963

; Sequence 11963, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 11963

; LENGTH: 732

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11963

Query Match 2.5%; Score 49.6; DB 4; Length 732;

Best Local Similarity 51.3%; Pred. No. 0.00059;

Matches 115; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```

Qy      591 GCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGA 650
        || ||  || |||  |  ||  |  ||  ||  |||  |||  ||  ||
Db      43  GCAGCGCTACGGCATGCCGCTGGAGCCTCGCCGGCTGGTCCATGGGCTGTCCATCGGCGA 102

Qy      651 GCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGA 710
        ||||  || |||  || |  ||  ||  ||  |  ||  ||  ||  ||
Db      103 GCGCCAGCGGGTGGAGATCGTGCCTGCCTGATGCAGGACATCCGCCTGCTGATCCTCGA 162

Qy      711 TGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCG 770
        || || ||||  | ||  |  ||  |  |||  |  |||  |  ||  ||  ||
Db      163 CGAGCCGACTTCGGTGCTGACCCACGCGAGGCCGAGGATCTCTTCGTCACCCTGCGCCG 222

Qy      771 CCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGC 814
        || ||  |  |  |  ||  ||  ||  |  |||  ||
Db      223 TCTTGCGGAAGAGGGCTGCAGTGTCTCTTCATCAGCCACAAGC 266

```

RESULT 13

US-09-252-991A-11890

; Sequence 11890, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 11890

; LENGTH: 2328

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11890

Query Match 2.5%; Score 49.6; DB 4; Length 2328;

Best Local Similarity 51.3%; Pred. No. 0.0011;

Matches 115; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```

Qy      591 GCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGA 650
          || || | || || | | | | | | | | | | | | | | | |
Db      45 GCAGCGCTACGGCATGCCGCTGGAGCCTCGCCGGCTGGTCCATGGGCTGTCCATCGGCGA 104

Qy      651 GCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGA 710
          || || | || || | | | | | | | | | | | | | | | |
Db      105 GCGCCAGCGGGTGGAGATCGTGCCTGCCTGATGCAGGACATCCGCCTGCTGATCCTCGA 164

Qy      711 TGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTCG 770
          || || || || | || | | | | | | | | | | | | | | |
Db      165 CGAGCCGACTTCGGTGCTGACCCACGCGAGGCCGAGGATCTCTTCGTACCCCTGCGCCG 224

Qy      771 CCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGC 814
          || || | | | | | | | | | | | | | | | | | | | |
Db      225 TCTTGCGGAAGAGGGCTGCAGTGTCTCTTCATCAGCCACAAGC 268

```

RESULT 14

US-09-252-991A-12050/c

; Sequence 12050, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12050
; LENGTH: 705
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12050

Query Match 2.4%; Score 49.4; DB 4; Length 705;
Best Local Similarity 54.0%; Pred. No. 0.00066;
Matches 101; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
Qy      628 GTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGG 687
          || | ||| |||| || ||||| || ||| || | || || |
Db      674 GTCCATGGGCTGTCCATCGGCGAGCGCCAGCGGGTGGAGATCGTGCGCTGCCTGATGCAG 615

Qy      688 AACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCAC 747
          || | | || || || || || || || || || || || || || || || ||
Db      614 GACATCCGCCTGCTGATCCTCGACGAGCCGACTTCGGTGCTGACCCACGCGAGGCCGAG 555

Qy      748 AATCTGGTGACAACCTTGTCGCCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTC 807
          |||| | ||| || ||| || || | | | | || || || | |
Db      554 GATCTCTTCGTCACCTGCGCCGTCTTGCGGAAGAGGGCTGCAGTGTCTCTTCATCAGC 495

Qy      808 CACCAGC 814
          ||| |||
Db      494 CACAAGC 488
```

RESULT 15

US-08-592-874-1/c

; Sequence 1, Application US/08592874
; Patent No. 5854034
; GENERAL INFORMATION:
; APPLICANT: POLLOCK, THOMAS J.
; APPLICANT: YAMAZAKI, MOTOHIDE
; APPLICANT: THORNE, LINDA
; APPLICANT: MIKOLAJCZAK, MARCIA
; APPLICANT: ARMENTROUT, RICHARD W.
; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
; TITLE OF INVENTION: POLYSACCHARIDE PRODUCTION
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JULES E. GOLDBERG
; STREET: 261 MADISON AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,874
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/377,440
; FILING DATE: 24-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDBERG, JULES E.
; REGISTRATION NUMBER: 24,408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-986-4090
; TELEFAX: 212-818-9479
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
US-08-592-874-1

```

```

Query Match          2.4%; Score 48.6; DB 2; Length 28804;
Best Local Similarity 46.6%; Pred. No. 0.0074;
Matches 156; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```

```

Qy      481 CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCAGGACCTTCTCC 540
      ||| | | | || || | ||||| | | || | | | | |||
Db      19377 CTGTTCAAGCGCTCGATCCGCGAGAACATTGCGCTGTCCAACCCGGCGATGCCGTTTCGAG 19318

Qy      541 CAGGCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC 600
      || | | | | | || | | | | | | | | | | | | |
Db      19317 CATGTCGTGGCGGCGGCGACGCTGGCGGGTGCGCATGACTTCATCCTGCGTCAGCCGCGC 19258

Qy      601 GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA 660
      | | | | | | | | | | | | | | | | | | | |
Db      19257 GGCTATGACACCGAGATCGTCGAGCGCGGCGTCAACCTGTGCGGCGGCCAGCGCCAGCGG 19198

Qy      661 GTGAGCATTGGGGTGCGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT 720
      | || | | || | | | | | | | | | | | | | |
Db      19197 CTCGCTATCGCCCGCGCGCTGGTTCGCAATCCGCGCATCCTGGTGTTCGACGAGGCGACC 19138

Qy      721 TCTGGCCTCGACAGCTTCACAGCCCAATCTGGTGACAACCTTGTCCCGCCTGGCCAAG 780
      || | | | | | | | | | | | | | | | | | | |
Db      19137 TCCGCGCTGGATGCCGAGAGCGAGGAGCTGATCCAGAACAATCTGCGCGCCATCTCGGCG 19078

Qy      781 GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCC 815
      ||| || ||||| |||| | | | | | | | |
Db      19077 GGCCGACGCTGGTGATCATCGCCACCGCCTGTC 19043

```

```

Search completed: February 26, 2004, 09:45:36
Job time : 114.675 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 06:40:42 ; Search time 463.956 Seconds
(without alignments)
15698.623 Million cell updates/sec

Title: US-09-989-981A-3
Perfect score: 2019
Sequence: 1 atggctgagaaaaccaaaga.....agtcaattcaagactgggtga 2019

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2353733 seqs, 1803733377 residues

Total number of hits satisfying chosen parameters: 4707466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							

1	2019	100.0	2019	10	US-09-989-981A-3	Sequence 3, Appli
2	1430	70.8	2669	10	US-09-989-981A-7	Sequence 7, Appli
3	743.8	36.8	3239	15	US-10-415-378-29	Sequence 29, Appl
4	199.2	9.9	2340	9	US-09-837-992-4	Sequence 4, Appli
5	199.2	9.9	2340	10	US-09-989-981A-5	Sequence 5, Appli
6	186.6	9.2	1959	10	US-09-989-981A-1	Sequence 1, Appli
7	186.6	9.2	2258	9	US-09-837-992-2	Sequence 2, Appli
8	169.2	8.4	2585	12	US-10-425-114-32175	Sequence 32175, A
9	137	6.8	2025	9	US-09-866-866A-13	Sequence 13, Appl
10	132.4	6.6	2027	15	US-10-405-806-1	Sequence 1, Appli
11	132.4	6.6	2053	15	US-10-405-806-12	Sequence 12, Appl
12	132.4	6.6	2247	9	US-09-866-866A-26	Sequence 26, Appl
13	132.4	6.6	2418	10	US-09-961-086-2	Sequence 2, Appli
14	132.4	6.6	2574	9	US-09-981-353-34	Sequence 34, Appl
15	132.4	6.6	2718	14	US-10-120-687-60	Sequence 60, Appl
16	132.4	6.6	2719	9	US-09-866-866A-9	Sequence 9, Appli
17	132.4	6.6	2883	14	US-10-101-510-639	Sequence 639, App
18	117.2	5.8	2930	9	US-09-954-531-591	Sequence 591, App
19	117.2	5.8	2930	14	US-10-171-581-276	Sequence 276, App
20	117.2	5.8	2930	16	US-10-429-160-9	Sequence 9, Appli
21	117	5.8	3201	13	US-10-072-621-5	Sequence 5, Appli
c 22	114.4	5.7	567	15	US-10-260-238-5734	Sequence 5734, Ap
23	114	5.6	2133	15	US-10-210-130-13	Sequence 13, Appl
24	110.2	5.5	925	12	US-10-424-599-64406	Sequence 64406, A
25	110.2	5.5	2028	15	US-10-369-493-27486	Sequence 27486, A
26	105.2	5.2	3463	12	US-10-425-114-13286	Sequence 13286, A
27	105.2	5.2	3528	12	US-10-424-599-99236	Sequence 99236, A
c 28	105	5.2	6043	10	US-09-989-981A-9	Sequence 9, Appli
29	104.6	5.2	2512	15	US-10-104-047-825	Sequence 825, App
30	104.4	5.2	2788	9	US-09-745-763-196	Sequence 196, App
31	102.8	5.1	972	12	US-10-424-599-129897	Sequence 129897,
c 32	101.4	5.0	447	9	US-09-960-352-11649	Sequence 11649, A
33	99.2	4.9	2826	12	US-10-424-599-95809	Sequence 95809, A
34	99	4.9	2638	15	US-10-369-493-27088	Sequence 27088, A
35	98.2	4.9	2223	9	US-09-938-842A-2262	Sequence 2262, Ap
36	98.2	4.9	2223	11	US-09-938-842A-2262	Sequence 2262, Ap
37	95.4	4.7	2400	13	US-10-108-605-244	Sequence 244, App
38	95	4.7	2608	12	US-10-424-599-33340	Sequence 33340, A
39	93.8	4.6	10330	13	US-10-001-189-68	Sequence 68, Appl
40	92	4.6	1684	12	US-10-425-114-4362	Sequence 4362, Ap
41	91.2	4.5	452	12	US-10-424-599-107241	Sequence 107241,
42	90.6	4.5	2546	12	US-10-424-599-33099	Sequence 33099, A
43	90.2	4.5	2162	12	US-10-425-114-25098	Sequence 25098, A
44	90.2	4.5	2248	12	US-10-425-114-31870	Sequence 31870, A
c 45	90	4.5	427	9	US-09-960-352-12839	Sequence 12839, A

ALIGNMENTS

RESULT 1

US-09-989-981A-3

; Sequence 3, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

```

; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2019)
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-3

```

```

Query Match          100.0%;  Score 2019;  DB 10;  Length 2019;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 2019;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCTGAGAAAACCAAAGAAGAGACCCAGCTGTGGAATGGGACTGTACTTCAGGATGCT 60

Qy     61 TCGGGCCTCCAGGACAGCTTGTTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCGGGCCTCCAGGACAGCTTGTTCCTCGGAAAGTGACAACAGTCTGTACTTCACCTAC 120

Qy    121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGTGGTCAGTCCAACACTCTGGAGGTCAGAGATCTCACCTACCAGGTGGACATCGCCTCT 180

Qy    181 CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGTGCCTTGGTTTGAGCAGCTGGCTCAGTTCAAGATACCCTGGAGGTCTCATAGCAGC 240

Qy    241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATG 300

Qy    301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGC 360

Qy    361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACG 420

```

Qy	421	CCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAAC	480
Db	421	CCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAAC	480
Qy	481	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCAGGACCTTCTCC	540
Db	481	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCAGGACCTTCTCC	540
Qy	541	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	541	CAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	660
Db	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	660
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Db	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Qy	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCCCTGGCCAAG	780
Db	721	TCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCCCTGGCCAAG	780
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC	960
Db	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC	960
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Db	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAG	1020
Qy	1021	AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1021	AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Qy	1081	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1081	TGGAAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Qy	1141	CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTTCCACC	1200
Db	1141	CAGGACACTGACTGTGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTTCCACC	1200
Qy	1201	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCATGGG	1260
Db	1201	CTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGCCCACGCTGCTCATTCATGGG	1260
Qy	1261	TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320

Db	1261		TCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCCAAG	1320
Qy	1321		CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Db	1321		CAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCTTTC	1380
Qy	1381		AATGTCATCCTGGATGTCGTCTCCAAATGTCACCTCGGAGAGGTCAATGCTGTACTATGAG	1440
Db	1381		AATGTCATCCTGGATGTCGTCTCCAAATGTCACCTCGGAGAGGTCAATGCTGTACTATGAG	1440
Qy	1441		CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Db	1441		CTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAATTG	1500
Qy	1501		CCGGAGCACTGTGCCTACGTATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Db	1501		CCGGAGCACTGTGCCTACGTATCATCTACGCGATGCCCATCTACTGGCTGACAAACCTG	1560
Qy	1561		CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620
Db	1561		CGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTCTGC	1620
Qy	1621		TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC	1680
Db	1621		TGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCACCTTCCACATGTCCTCCTTC	1680
Qy	1681		TTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC	1740
Db	1681		TTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTGGAC	1740
Qy	1741		AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTCTCG	1800
Db	1741		AACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGCTTCTCG	1800
Qy	1801		GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC	1860
Db	1801		GGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTCACC	1860
Qy	1861		TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT	1920
Db	1861		TTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTCTAT	1920
Qy	1921		GCGATCTACCTCATTTGTCATCGGCATCAGCTACGGCTTCCTGTTCCCTGTACTATCTATCC	1980
Db	1921		GCGATCTACCTCATTTGTCATCGGCATCAGCTACGGCTTCCTGTTCCCTGTACTATCTATCC	1980
Qy	1981		TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019
Db	1981		TTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA	2019

RESULT 2

US-09-989-981A-7

; Sequence 7, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

Qy	421	CCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAAC	480
Db	520		579
Qy	481	CTGACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCAGGACCTTCTCC	540
Db	580		639
Qy	541	CAGGCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGC	600
Db	640		699
Qy	601	GCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGA	660
Db	700		759
Qy	661	GTGAGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACT	720
Db	760		819
Qy	721	TCTGGCCTCGACAGCTTCACAGCCACAATCTGGTGACAACCTTGTCGCCGCTGGCCAAG	780
Db	820		879
Qy	781	GGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTT	840
Db	880		939
Qy	841	GACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATG	900
Db	940		999
Qy	901	GTGCAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTC	960
Db	1000		1059
Qy	961	TACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGAGGTGGCCACCGTGGAG	1020
Db	1060		1119
Qy	1021	AAGGCACAGTCTCTTGACGCCCTGTTCTAGAAAAAGTACAAGGCTTTGATGACTTTCTG	1080
Db	1120		1179
Qy	1081	TGGAAAGCTGAGGCAAAGGAACCAACACAAGCACCCACACAGTCAGCCTGACCCTCACA	1140
Db	1180		1239
Qy	1141	CAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCCGGGATGATAGAGCAGTTTCTC	1197
Db	1240		1299
Qy	1198	ACCCTGATCCGTCGTGAGATTTCCAATGACTTCCGGGACCTGCCACGCTGCTCATTCAT	1257
Db	1300		1359

Qy 1258 GGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTTACTACGGCCATGGGGCC 1317
 ||| ||| ||| ||| ||| ||| | | ||| ||| ||| || | ||| ||| |
 Db 1360 GGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCTATTTTGGCCATGGGAGC 1419

Qy 1318 AAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGATAGGGGCGCTCATTCCT 1377
 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1420 ATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTCATGATCGGTGCTCTCATCCCT 1479

Qy 1378 TTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACTCGGAGAGGTCAATGCTGTACTAT 1437
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1480 TTCAACGTCATTCTGGATGTCATCTCCAAATGTTACTCAGAGAGGGCAATGCTTTACTAT 1539

Qy 1438 GAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTGCCAAGATCCTAGGAGAA 1497
 || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1540 GAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTGCCAAGATCCTCGGGGAG 1599

Qy 1498 TTGCCGGAGCACTGTGCCTACGTCATCATCTACGCGATGCCCATCTACTGGCTGACAAAC 1557
 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1600 CTTCCGGAGCACTGTGCCTACATCATCATCTACGGGATGCCACCTACTGGCTGGCCAAC 1659

Qy 1558 CTGCGGCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCGTGTGGTTGGTGGTCTTC 1617
 ||| ||| | | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1660 CTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGGTGTGGCTGGTGGTCTTC 1719

Qy 1618 TGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCACCTTCCACATGTCCTCC 1677
 || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1720 TGTTCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCACCTTCCACATGGCCTCC 1779

Qy 1678 TTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCGGCTTCATGATAAACTTG 1737
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1780 TTCTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGGGGCTTCATGATAAACTTG 1839

Qy 1738 GACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGTTCCCTCCGGTGGTGGTTC 1797
 || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1840 AGCAGCCTGTGGACAGTGCCCGCTGGATTTCCAAGTGTCTTCCTGCGGTGGTGTTC 1899

Qy 1798 TCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCACACAAATCGGCAACTTC 1857
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1900 GAAGGGCTGATGAAGATTCAGTTTCAGCAGAAGAACTTATAAAATGCCTCTCGGGAACCTC 1959

Qy 1858 ACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACCTGAACTCGCATCCACTC 1917
 ||| || | || ||| || || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1960 ACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGCTGGACTCGTACCCTCTC 2019

Qy 1918 TATGCGATCTACCTCATTGTATCGGCATCAGCTACGGCTTCCTGTTCTGTACTATCTA 1977
 || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 2020 TACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCATGGTCCTGTACTACGTG 2079

Qy 1978 TCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
 ||| || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 2080 TCCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA 2121

RESULT 3
 US-10-415-378-29
 ; Sequence 29, Application US/10415378
 ; Publication No. US20040014945A1

```

; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom
; APPLICANT: YUE, Henry; NGUYEN, Dannel B.;
; APPLICANT: HAFALIA, April J.A.; ELLIOTT, Vicki S.;
; APPLICANT: LU, Yan; CHAWLA, Narinder K.;
; APPLICANT: YAO, Monique G.; BAUGHN, Mariah R.;
; APPLICANT: GANDHI, Ameena R.; DING, Li;
; APPLICANT: SANJANWALA, Madhusudan M.; RAMKUMAR, Jayalaxmi;
; APPLICANT: ARVIZU, Chandra S.; GIETZEN, Kimberly J.;
; APPLICANT: LAL, Preeti G.; AZIMZAI, Yalda;
; APPLICANT: KHAN, Farrah A.; THANGAVELU, Kavitha;
; APPLICANT: THORNTON, Michael B.; LU, Dyung Aina M.;
; APPLICANT: TRIBOULEY, Catherine M.; WARREN, Bridget A.;
; APPLICANT: ISON, H. Craig; DAS, Debopriya;
; APPLICANT: RAUMANN, Brigitte E.; POLICKY, Jennifer L.;
; APPLICANT: KEARNEY, Liam
; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
; FILE REFERENCE: PI-0270 USN
; CURRENT APPLICATION NUMBER: US/10/415,378
; CURRENT FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: PCT/US01/46055
; PRIOR FILING DATE: 2001-10-27
; PRIOR APPLICATION NUMBER: US 60/250,790
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/252,232
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/249,661
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 60/247,673
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/245,904
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/243,989
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 3239
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040014945A1 6585710CB1
US-10-415-378-29

```

```

Query Match          36.8%; Score 743.8; DB 15; Length 3239;
Best Local Similarity 78.9%; Pred. No. 7.1e-221;
Matches 899; Conservative 0; Mismatches 237; Indels 3; Gaps 1;

```

```

Qy      884 GGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATA 943
      ||| | |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      12 GGGGCGGCCAGCACATGGTCCATTATTTACAGCCATCGGCTACCCCTGTCTCGCTACA 71

Qy      944 GCAACCCTGCGGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGG 1003
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||| | |||
Db      72 GCAATCCTGCTGACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGG 131

```

Qy 1004 AGGTGGCCACCGTGGAGAAGGCACAGTCTCTTGAGCCCTGTTCTAGAAAAAGTACAAG 1063
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 132 AATTGGCCACCAGGGAGAAGGCTCAGTCACTCGCAGCCCTGTTTCTAGAAAAAGTGCCTG 191

Qy 1064 GCTTTGATGACTTTCTGTGGAAGCTGAGGCAAAGGAACTCAACACAAGCACCCACACAG 1123
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 192 ACTTAGATGACTTTCTATGGAAAGCAGAGACGAAGGATCTTGACGAGGACACCTGTGTGG 251

Qy 1124 TCAGCCTGACCCTCACACAGGACACTGACTG---TGGGACTGCTGTTGAGCTGCCCCGGA 1180
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 252 AAAGCAGCGTGACCCCACTAGACACCAACTGCCTCCCGAGTCCTACGAAGATGCCTGGGG 311

Qy 1181 TGATAGAGCAGTTTTCCACCCTGATCCGTCGTCAGATTTCCAATGACTTCCGGGACCTGC 1240
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 312 CGGTGCAGCAGTTTACGACGCTGATCCGTCGTCAGATTTCCAACGACTTCCGAGACCTGC 371

Qy 1241 CCACGCTGCTCATTCATGGGTCGGAAGCCTGCCTGATGTCCCTCATCATTGGCTTCCTTT 1300
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 372 CCACCCTCCTCATCCATGGGGCGGAGGCCTGTCTGATGTCAATGACCATCGGCTTCCTCT 431

Qy 1301 ACTACGGCCATGGGGCCAAGCAGCTCTCCTTCATGGACACAGCAGCCCTCCTCTTCATGA 1360
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 432 ATTTTGGCCATGGGAGCATCCAGCTCTCCTTCATGGATACAGCCGCCCTCTTGTTTCATGA 491

Qy 1361 TAGGGGCGCTCATTCCTTTCAATGTCATCCTGGATGTCGTCTCCAAATGTCACCTCGGAGA 1420
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 492 TCGGTGCTCTCATCCCTTTCAACGTCAATTCTGGATGTCATCTCCAAATGTTACTCAGAGA 551

Qy 1421 GGTCAATGCTGTACTATGAGCTGGAAGACGGGCTGTACACTGCTGGTCCTTATTTCTTTG 1480
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 552 GGGCAATGCTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTTG 611

Qy 1481 CCAAGATCCTAGGAGAATTGCCGGAGCACTGTGCCTACGTCATCATCTACGGATGCCCA 1540
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 612 CCAAGATCCTCGGGGAGCTTCCGGAGCACTGTGCCTACATCATCATCTACGGGATGCCCA 671

Qy 1541 TCTACTGGCTGACAAACCTGCGGCCCCGTGCCTGAGCTCTTCCTTCTACACTTCCTGCTCG 1600
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 672 CCTACTGGCTGGCCAACCTGAGGCCAGGCCTCCAGCCCTTCCTGCTGCACTTCCTGCTGG 731

Qy 1601 TGTGGTTGGTGGTCTTCTGCTGCAGGACCATGGCCCTGGCTGCCTCTGCCATGCTGCCCA 1660
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 732 TGTGGCTGGTGGTCTTCTGTTGCAGGATTATGGCCCTGGCCGCCGCGGCCCTGCTCCCCA 791

Qy 1661 CCTTCCACATGTCCTCCTTCTTCTGCAATGCCCTCTACAACCTCCTTCTACCTTACTGCCG 1720
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 792 CCTTCCACATGGCCTCCTTCTTTCAGCAATGCCCTCTACAACCTCCTTCTACCTCGCCGGG 851

Qy 1721 GCTTCATGATAAACTTGGACAACCTGTGGATAGTGCCTGCATGGATCTCCAAGCTGTCGT 1780
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 852 GCTTCATGATAAACTTGAGCAGCCTGTGGACAGTGCCCGCGTGGATTTCCAAAGTGTCT 911

Qy 1781 TCCTCCGGTGGTGGTCTTCTCGGGGCTGATGCAGATTCAATTTAATGGACACCTTTACACCA 1840
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 912 TCCTGCGGTGGTGT'TTGAAGGGCTGATGAAGATTCAAGTTCAGCAGAAGAACTTATAAAA 971

Qy 1841 CACAAATCGGCAACTTCACCTTCTCCATCCTCGGAGACACGATGATCAGTGCCATGGACC 1900

```

      |  ||| ||| |||| | | |  ||  |||| |  ||  ||||| ||||| |
Db      972 TGCCTCTCGGGAACCTCACCATCGCGGTCTCAGGAGATAAAATCCTCAGTGCCATGGAGC 1031

Qy      1901 TGAACTCGCATCCACTCTATGCGATCTACCTCATTGTCATCGGCATCAGCTACGGCTTCC 1960
      || |||| |  || |||| |  ||||| ||||| ||||| ||| ||||  |||||
Db      1032 TGGACTCGTACCCTCTCTACGCCATCTACCTCATCGTCATTGGCCTCAGCGGTGGCTTCA 1091

Qy      1961 TGTTCTGTACTATCTATCCTTGAAGCTCATCAAACAGAAGTCAATTCAAGACTGGTGA 2019
      || ||||| |||| |  |||| |  ||||| ||||| ||| ||||| |||||
Db      1092 TGGTCCTGTACTACGTGTCTTAAGGTTTCATCAAACAGAAACCAAGTCAAGACTGGTGA 1150

```

RESULT 4

US-09-837-992-4

; Sequence 4, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 2340

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: human sitosterolemia gene (SSG)

; NAME/KEY: CDS

; LOCATION: (107)..(2062)

; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)

; OTHER INFORMATION: protein

US-09-837-992-4

Query Match 9.9%; Score 199.2; DB 9; Length 2340;

Best Local Similarity 54.0%; Pred. No. 6.1e-51;

Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

```

Qy      234 TAGCAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGG 293
      | || | || |  || |  || |  || |  || |  || |  || |  || |
Db      283 TTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGG 342

Qy      294 ACAGATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGAT 353
      |||| |  || ||| |||| ||||| ||||| |||| |  || |  || |  || |  ||
Db      343 GCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCAT 402

Qy      354 CACAGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACC 413

```

```

      | | | | |      | |      | |      | | | | |      | | | | |
Db      403 GTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGC 462

Qy      414 CAGTACGCCTCAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCT 473
      | | | | | | | | | | | | | | | | | | | | | |
Db      463 GCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCT 522

Qy      474 GCCCAACCTGACCGTCAGAGAGACCCTGGCTTTCATTGCCAGATGCGCCTGCCAGGAC 533
      | | | | | | | | | | | | | | | | | | | | | |
Db      523 GAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGG 582

Qy      534 CTTCTCCCAGGCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCG 593
      | | | | | | | | | | | | | | | | | | | | | |
Db      583 CAATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAG 639

Qy      594 GCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCGGGGGTGAGCG 653
      | | | | | | | | | | | | | | | | | | | | | |
Db      640 CCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCG 699

Qy      654 CCGACGAGTGAGCATTGGGGTGCGAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGA 713
      | | | | | | | | | | | | | | | | | | | | | |
Db      700 GCGCCGGGTCTCCATCGCAGCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGA 759

Qy      714 ACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCCGCT 773
      | | | | | | | | | | | | | | | | | | | | | |
Db      760 GCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAC 819

Qy      774 GGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAG 833
      | | | | | | | | | | | | | | | | | | | | | |
Db      820 GGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACAGCCCCGTTCTGAGCTTTTCA 879

Qy      834 GCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCA 893
      | | | | | | | | | | | | | | | | | | | | | |
Db      880 GCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGC 939

Qy      894 GCAAATGGTGCAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGC 953
      | | | | | | | | | | | | | | | | | | | | | |
Db      940 GGAAATGCTTGATTTCTTCAATGACTGCGGTTACCCTTGTCCTGAACATTCAAACCCTTT 999

Qy      954 GGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCAC 1013
      | | | | | | | | | | | | | | | | | | | | | |
Db      1000 TGACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAAC 1059

Qy      1014 CGTGGAGAAGGCACAG 1029
      | | | | | | | | | | | | | | | | | | | | | |
Db      1060 CTCCAAGAGAGTCCAG 1075

```

RESULT 5

US-09-989-981A-5

; Sequence 5, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui


```

; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-5

```

```

Query Match          9.9%; Score 199.2; DB 10; Length 2340;
Best Local Similarity 54.0%; Pred. No. 6.1e-51;
Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

```

```

Qy      234 TAGCAGCCAAGACTCCTGTGAGCTGGGCATCCGAAATCTAAGCTTCAAAGTGAGGAGTGG 293
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      283 TTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGG 342

Qy      294 ACAGATGCTGGCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGAT 353
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      343 GCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCAT 402

Qy      354 CACAGGCAGAGGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACC 413
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      403 GTCCGGGAGGCTGGGGCGCGCGGGGACCTTCTCTGGGGGAGGTGTATGTGAACGGCCGGGC 462

Qy      414 CAGTACGCCTCAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCT 473
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      463 GCTGCGCCGGGAGCAGTTCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCT 522

Qy      474 GCCCAACCTGACCGTCAGAGAGACCCTGGCTTTTCATTGCCAGATGCGCCTGCCAGGAC 533
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      523 GAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGG 582

Qy      534 CTTCTCCCAGGCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCG 593
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      583 CAATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAG 639

Qy      594 GCAGTGCGCCAACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCG 653
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      640 CCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCG 699

Qy      654 CCGACGAGTGAGCATTGGGGTGACGCTCCTGTGGAACCCAGGAATCCTCATTTCTGGATGA 713
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      700 GCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGA 759

```

Qy 714 ACCCACTTCTGGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCT 773
 || || | |||| ||| || | || || | | || | || ||
 Db 760 GCCAACCAAGGCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAC 819
 Qy 774 GGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAG 833
 ||| | ||| | | ||| | || | | ||||| || |||| | ||
 Db 820 GGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTCA 879
 Qy 834 GCTATTTGACCTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCA 893
 ||| ||||| | | ||| | || | | || | | || ||
 Db 880 GCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGC 939
 Qy 894 GCAAATGGTGCAGTACTTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGC 953
 | |||| | | |||| | || | ||||| || |||||
 Db 940 GGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTT 999
 Qy 954 GGACTTCTACGTGGACTTGACCAGCATCGACAGACGCAGCAAAGAACGGGAGGTGGCCAC 1013
 ||||| |||| ||| | || | | |||| ||||| | | ||
 Db 1000 TGACTTCTATATGGACCTGACGTCAGTGGATAACCCAAAGCAAGGAACGGGAAATAGAAAC 1059
 Qy 1014 CGTGGAGAAGGCACAG 1029
 | || | |||
 Db 1060 CTCCAAGAGAGTCCAG 1075

RESULT 6

US-09-989-981A-1

; Sequence 1, Application US/09989981A
 ; Publication No. US20030049730A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hobbs, Helen H.
 ; APPLICANT: Shan, Bei
 ; APPLICANT: Barnes, Robert
 ; APPLICANT: Tian, Hui
 ; APPLICANT: Tularik Inc.
 ; APPLICANT: Board of Regents, The University of Texas System
 ; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
 ; FILE REFERENCE: 018781-007320US
 ; CURRENT APPLICATION NUMBER: US/09/989,981A
 ; CURRENT FILING DATE: 2002-07-23
 ; PRIOR APPLICATION NUMBER: US 60/252,235
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/253,645
 ; PRIOR FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 1959
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1959)
 ; OTHER INFORMATION: mouse ABCG5 (mABCG5)
 US-09-989-981A-1

Query Match 9.2%; Score 186.6; DB 10; Length 1959;
 Best Local Similarity 53.1%; Pred. No. 4.8e-47;
 Matches 421; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

Qy	261	CATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTC	320
Db	207	CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTC	266
Qy	321	AGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGAT	380
Db	267	AGGCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGAC	326
Qy	381	GAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGTG	440
Db	327	CCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTG	386
Qy	441	CGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCAACCTGACCGTCAGAGAGACCCT	500
Db	387	CTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTT	446
Qy	501	GGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCCAGCGTGACAAACG	560
Db	447	GCGATACACAGC---GATGCTGGCCCTCTGCCCGAGCTCCGCGGACTTCTACAACAAGAA	503
Qy	561	GGTGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAA	620
Db	504	GGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAATGATTGGCAG	563
Qy	621	CACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCT	680
Db	564	CTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGCAGCCCACT	623
Qy	681	CCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCAC	740
Db	624	CCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGAC	683
Qy	741	AGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGCAACAGGCTGGTGCTCAT	800
Db	684	TGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGT	743
Qy	801	CTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGAC	860
Db	744	CACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGAC	803
Qy	861	ATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCACTACTTCACATCCAT	920
Db	804	TTACGGAGAGTTGGTGTCTGTGGCAGCCAGAGGAGATGCTTGGCTTCTTCAATAACTG	863
Qy	921	TGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCAT	980
Db	864	TGGTTACCCCTGTCTGAACATTCCAATCCCTTTGATTTTACATGGACTTGACATCAGT	923
Qy	981	CGACAGACGCAGCAAAGAACGGGAGGTGGCCACCGTGGAGAAGGCACAGTCTCTTGCAGC	1040
Db	924	GGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTACAAGCGAGTACAGATGCTGGAATG	983
Qy	1041	CCTGTTCTTAGAA	1053

Db ||| |||
984 TGCCTTCAAGGAA 996

RESULT 7

US-09-837-992-2

; Sequence 2, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2258

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)

; NAME/KEY: CDS

; LOCATION: (47)..(2005)

; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)

; OTHER INFORMATION: protein

US-09-837-992-2

Query Match 9.2%; Score 186.6; DB 9; Length 2258;

Best Local Similarity 53.1%; Pred. No. 5.1e-47;

Matches 421; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

```
Qy      261 CATCCGAAATCTAAGCTTCAAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTC 320
          | || | || | || | || | || | || | || | || | || | || | || |
Db      253 CCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTC 312

Qy      321 AGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGAT 380
          |||| | || | || | || | || | || | || | || | || | || |
Db      313 AGGCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGAC 372

Qy      381 GAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGTG 440
          ||| | | | | || | || | || | || | || | || |
Db      373 CCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCACTTCCAAGACTG 432

Qy      441 CGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCCT 500
          | | | | | || | || | || | || | || | || | || | || |
Db      433 CTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTT 492

Qy      501 GGCTTTCATTGCCAGATGCGCCTGCCAGACCTTCTCCAGGCCCAGCGTGACAAACG 560
```


; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 32175
 ; LENGTH: 2585
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: UC-ZMFLB73274A02_FLI
 US-10-425-114-32175

Query Match 8.4%; Score 169.2; DB 12; Length 2585;
 Best Local Similarity 53.4%; Pred. No. 1.5e-41;
 Matches 382; Conservative 0; Mismatches 328; Indels 6; Gaps 1;

Qy	269	ATCTAAGCTTCAAAGTGAGGAGTGGACAGATGCTGGCCATCATAGGGAGCTCAGGCTGCG	328
Db	582	AGCTCACCGGGTACGCGGAGCCCGGTCGCTGACCGCGCTCATGGGGCCCTCGGGGTCCG	641
Qy	329	GGAGAGCCTCACTACTCGACGTGATCACAGGCAGAGGCCACGGTGGCAAGATGAAATCAG	388
Db	642	GCAAGTCCACCCTGCTCGACGCCCTCGCCGGCCGCTCGCCGCCAACGCCTTCCTCTCCG	701
Qy	389	GACAAATTTGGATAAATGGGCAACCCAGTACGCCTCAGCTGGTGAGGAAGTGCGTTGCGC	448
Db	702	GCAACGTGCTCCTCAACGG-----CCGCAAGGCCAAGCTCTCCTTCGGCGCCGCGGCGT	755
Qy	449	ATGTGCGGCAGCATGACCAACTGCTGCCCAACCTGACCGTCAGAGAGACCCTGGCTTTCA	508
Db	756	ACGTGACGCAGGACGACAACCTGATCGGGACGCTGACGGTGCGCGAGACGATCGGCTACT	815
Qy	509	TTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAGGCCAGCGTGACAAACGGGTGGAAG	568
Db	816	CGGCGCTGCTGCGGCTGCCGACAAGATGCCGCGGGAGGACAAGCGCGCGCTGGTGGAGG	875
Qy	569	ACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCCAACACCAGAGTGGGCAACACGTATG	628
Db	876	GCACCATCGTCGAGATGGGGCTGCAGGACTGCGCCGACACCGTCATCGGCAACTGGCACC	935
Qy	629	TACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTGAGCATTGGGGTGCAGCTCCTGTGGA	688
Db	936	TCCGCGGGGTACGCGCGGCGAGAAGCGCCGCGTCAGCATCGCGCTCGAGCTACTCATGC	995
Qy	689	ACCCAGGAATCCTCATTTCTGGATGAACCCACTTCTGGCCTCGACAGCTTCACAGCCCACA	748
Db	996	GCCCGCGCCTCCTCTTCTCGACGAGCCACCAGCGGCCTCGACAGCTCGTCTGCGTTCT	1055
Qy	749	ATCTGGTGACAACCTTGTCCC GCCTGGCCAAGGGCAACAGGCTGGTGCTCATCTCCCTCC	808
Db	1056	TCGTGACGCAGACGCTGCGGGGCCTGGCGAGGGACGGCAGGACGGTGATTGCTTCCATCC	1115
Qy	809	ACCAGCCTCGCTCTGACATCTTCAGGCTATTTGACCTGGTCCTTCTGATGACATCTGGCA	868
Db	1116	ACCAGCCCAGCAGCGAGGTGTTTCGAGCTCTTCGACATGCTCTTCCTGCTATCCGGGGGCA	1175
Qy	869	CCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTGCAGTACTTCACATCCATTGGCCACC	928
Db	1176	AGACCGTCTACTTCGGACAAGCATGCAAGCATGCGAGTTCTTTGCTCAAGCCGGTTTCC	1235

Qy 929 CTTGTCTCGCTATAGCAACCCTGCGGACTTCTACGTGGACTTGACCAGCATCGAC 984
 |||| || | || | || | || | || | ||
 Db 1236 CTTGCCCCGGCTCTGCGGAATCCGTCCGACCATTTCCTGAGGTGCGTCAACTCGGAC 1291

Query Match 6.8%; Score 137; DB 9; Length 2025;
Best Local Similarity 52.3%; Pred. No. 1.5e-31;
Matches 352; Conservative 0; Mismatches 315; Indels 6; Gaps 2;

Qy 664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
 ||||| ||| || ||||| | ||||| ||||| ||||| ||
 Db 599 AGCATAGGGATGGAGCTGATCACTGACCCTTCCATCCTCTTCTGGATGAGCCACGACT 658
 Qy 724 GGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTCGCCTGGCCAAGGGC 783
 || | ||| ||||| | | | | | || | || | ||
 Db 659 GGTTCGACTCAAGCACAGCGAATGCTGTCCTTTTGCTCCTGAAAAGGATGTCTAAACAG 718
 Qy 784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
 | | | ||||| | | ||||| | | ||||| | | |||||
 Db 719 GGTGCAACAATCATCTTCTCCATTTCATCAGCCTCGGTATTCATCTTTAAGTTGTTTGAC 778
 Qy 844 CTGGTCCTTCTGATGACATCTGGCACCCTATCTACCTGGGGGCGGCGCAGCAATGGTG 903
 || | || | || || | | | | | | | | | | | | |
 Db 779 AGCCTCACCTTACTGGCTTCCGGGAAACTCGTGTTCCATGGGCCAGCACAGAAGGCCTTG 838
 Qy 904 CAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTAC 963
 ||||| |||| | || || || | || | ||||| || | |
 Db 839 GAGTACTTTGCATCAGCAGGTTACCACTGTGAGCCCTACAACAACCCTGCGGATTTTTTC 898
 Qy 964 GTGGACTTGACCA 976
 | || | ||
 Db 899 CTTGATGTCATCA 911

RESULT 10

US-10-405-806-1

; Sequence 1, Application US/10405806

; Publication No. US20030232362A1

; GENERAL INFORMATION:

; APPLICANT: KOMATANI, HIDEYA

; APPLICANT: HARA, YOSHIKAZU

; APPLICANT: KOTANI, HIDEHITO

; APPLICANT: NAKAGAWA, RINAKO

; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF

; FILE REFERENCE: 234985USOCONT

; CURRENT APPLICATION NUMBER: US/10/405,806

; CURRENT FILING DATE: 2003-04-03

; PRIOR APPLICATION NUMBER: PCT/JP01/08112

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: JP2000-303441

; PRIOR FILING DATE: 2000-10-03

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1

; LENGTH: 2027

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (45)..(2009)

US-10-405-806-1

Query Match 6.6%; Score 132.4; DB 15; Length 2027;

Best Local Similarity 51.9%; Pred. No. 4e-30;

Matches 348; Conservative 0; Mismatches 316; Indels 6; Gaps 2;

Qy 304 GCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGA 363
 ||||| | | | | | | | | | | | | | | | | | |
 Db 273 GCCATCCTGGGACCCACAGGTGGAGGCAAATCTTCGTTATTAGATG---TCTTAGCTGCA 329
 Qy 364 GGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCT 423
 | | | | | | | | | | | | | | | | | |
 Db 330 AGGAAAGATCCAAGTGGATTATCTGGAGATGTTCTGATAAATGGAGCACCGCGACCTGCC 389
 Qy 424 CAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTG 483
 | | | | | | | | | | | | | | | | | |
 Db 390 AATTTCAAATGTAATTCAGGT---TACGTGGTACAAGATGATGTTGTGATGGGCACTCTG 446
 Qy 484 ACCGTCAGAGAGACCCTGGCTTTCATTGCCAGATGCGCCTGCCCAGGACCTTCTCCAG 543
 || || ||||| | | | | | | | | | | | | | |
 Db 447 ACGGTGAGAGAAAACCTTACAGTTCTCAGCAGCTCTTCGGCTTGCAACAACTATGACGAAT 506
 Qy 544 GCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCC 603
 | | | | | | | | | | | | | | | | | |
 Db 507 CATGAAAAAACGAACGGATTAACAGGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGA 566
 Qy 604 AACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTG 663
 || ||| || || | | | | | | | | | | | | | |
 Db 567 GACTCCAAGGTTGGAACCTCAGTTTATCCGTGGTGTGTCTGGAGGAGAAAGAAAAGGACT 626
 Qy 664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
 || || | | |||| | | | | | | | | | | | |
 Db 627 AGTATAGGAATGGAGCTTATCACTGATCCTTCATCTTGTCTTGGATGAGCCTACAAC 686
 Qy 724 GGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCGCCGCTGGCCAAGGGC 783
 ||| | ||| ||||| | | | | | | | | | | | |
 Db 687 GGCTTAGACTCAAGCACAGCAAATGCTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAG 746
 Qy 784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
 | | | | | | | | | | | | | | | | | | | |
 Db 747 GGACGAACAATCATCTTCTCCATTTCATCAGCCTCGATATTCATCTTCAAGTTGTTTGAT 806
 Qy 844 CTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTG 903
 || | | | | | | | | | | | | | | | | | |
 Db 807 AGCCTCACCTTATTGGCCTCAGGAAGACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTG 866
 Qy 904 CAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTAC 963
 |||| | | | | | | | | | | | | | | | |
 Db 867 GGATACTTTGAATCAGCTGGTTATCACTGTGAGGCCTATAATAACCCTGCAGACTTCTTC 926
 Qy 964 GTGGACTTGA 973
 |||| | |
 Db 927 TTGGACATCA 936

RESULT 11

US-10-405-806-12

; Sequence 12, Application US/10405806

; Publication No. US20030232362A1

; GENERAL INFORMATION:

; APPLICANT: KOMATANI, HIDEYA

; APPLICANT: HARA, YOSHIKAZU

```
; APPLICANT: KOTANI, HIDEHITO
; APPLICANT: NAKAGAWA, RINAKO
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF
; FILE REFERENCE: 234985US0CONT
; CURRENT APPLICATION NUMBER: US/10/405,806
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: PCT/JP01/08112
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP2000-303441
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 2053
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ABCG2 482Tmutant sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32)..(1999)
US-10-405-806-12
```

```
Query Match          6.6%; Score 132.4; DB 15; Length 2053;
Best Local Similarity 51.9%; Pred. No. 4.1e-30;
Matches 348; Conservative 0; Mismatches 316; Indels 6; Gaps 2;
```

```
Qy      304 GCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGA 363
      ||||| | || | |||| | || | | || | || | || | || |
Db      260 GCCATCCTGGGACCCACAGGTGGAGGCAAATCTTCGTTATTAGATG---TCTTAGCTGCA 316

Qy      364 GGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCT 423
      | | | | | ||||| | || | ||||| || | | | |
Db      317 AGGAAAGATCCAAGTGGATTATCTGGAGATGTTCTGATAAATGGAGCACCGCGACCTGCC 376

Qy      424 CAGCTGGTGAGGAAGTGC GTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTG 483
      | | | | | ||||| || ||||| || || || || ||
Db      377 AATTTCAAATGTAATTCAGGT---TACGTGGTACAAGATGATGTTGTGATGGGCACTCTG 433

Qy      484 ACCGTCAGAGAGACCCTGGCTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAG 543
      || || ||||| | | | |||| | | || || | | | |
Db      434 ACGGTGAGAGAAAAC TTACAGTTCTCAGCAGCTCTTCGGCTTGCAACAAC TATGACGAAT 493

Qy      544 GCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCC 603
      | | ||||| | | |||| | || || || || | | ||
Db      494 CATGAAAAAACGAACGGATTAAACAGGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCA 553

Qy      604 AACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTG 663
      || ||| || || | | | | ||||| ||||| || || || | |
Db      554 GACTCCAAGGTTGGAAC TCAGTTTATCCGTGGTGTGTCTGGAGGAGAAAGAAAAGGACT 613

Qy      664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
      || || || || |||| | | || || || || || || || ||
Db      614 AGTATAGGAATGGAGCTTATCACTGATCCTTCCATCTTGTTCTTGGATGAGCCTACAAC T 673

Qy      724 GGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGC 783
      ||| | ||| ||||| | | | | || | | || | |||
```

Db 674 GGCTTAGACTCAAGCACAGCAAATGCTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAG 733
 QY 784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
 | | | | | | | | | | | | | | | | | | | | | |
 Db 734 GGACGAACAATCATCTTCTCCATTTCATCAGCCTCGATATTCATCTTCAAGTTGTTTGAT 793
 QY 844 CTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTG 903
 | | | | | | | | | | | | | | | | | | | | | |
 Db 794 AGCCTCACCTTATTGGCCTCAGGAAGACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTG 853
 QY 904 CAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTAC 963
 | | | | | | | | | | | | | | | | | | | | | |
 Db 854 GGATACTTTGAATCAGCTGTTATCACTGTGAGGCCTATAATAACCCTGCAGACTTCTTC 913
 QY 964 GTGGACTTGA 973
 | | | | | | |
 Db 914 TTGGACATCA 923

RESULT 12

US-09-866-866A-26

; Sequence 26, Application US/09866866A

; Patent No. US20020102244A1

; GENERAL INFORMATION:

; APPLICANT: Sorrentino, Brian

; APPLICANT: Schuetz, John

; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells

; FILE REFERENCE: 1340-1-021CIP2

; CURRENT APPLICATION NUMBER: US/09/866,866A

; CURRENT FILING DATE: 2001-08-30

; PRIOR APPLICATION NUMBER: 09/584,586

; PRIOR FILING DATE: 2000-05-31

; PRIOR APPLICATION NUMBER: PCT/US99/11825

; PRIOR FILING DATE: 1999-05-27

; PRIOR APPLICATION NUMBER: 60/086,988

; PRIOR FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 26

; LENGTH: 2247

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-866-866A-26

Query Match 6.6%; Score 132.4; DB 9; Length 2247;
 Best Local Similarity 51.9%; Pred. No. 4.2e-30;
 Matches 348; Conservative 0; Mismatches 316; Indels 6; Gaps 2;

QY 304 GCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGA 363
 | | | | | | | | | | | | | | | | | | | | | |
 Db 422 GCCATCCTGGGACCCACAGGTGGAGGCAAATCTTCGTTATTAGATG---TCTTAGCTGCA 478
 QY 364 GGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCT 423
 | | | | | | | | | | | | | | | | | | | | | |
 Db 479 AGGAAAGATCCAAGTGGATTATCTGGAGATGTTCTGATAAATGGAGCACCGCGACCTGCC 538
 QY 424 CAGCTGGTGAGGAAGTGC GTTGC GCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTG 483

```

      | | | | | | | | | | | | | | | | | | | | | |
Db      539 AATTTCAAATGTAATTCAGGT---TACGTGGTACAAGATGATGTTGTGATGGGCACTCTG 595

Qy      484 ACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCAGGACCTTCTCCCAG 543
      | | | | | | | | | | | | | | | | | | | | | |
Db      596 ACGGTGAGAGAAAACCTTACAGTTCTCAGCAGCTCTTCGGCTTGCAACAACATATGACGAAT 655

Qy      544 GCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCC 603
      | | | | | | | | | | | | | | | | | | | | | |
Db      656 CATGAAAAAACGAACGGATTAACAGGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCA 715

Qy      604 AACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTG 663
      | | | | | | | | | | | | | | | | | | | | | |
Db      716 GACTCCAAGGTTGGAACCTCAGTTTATCCGTGGTGTGTCTGGAGGAGAAAGAAAAAGGACT 775

Qy      664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
      | | | | | | | | | | | | | | | | | | | | | |
Db      776 AGTATAGGAATGGAGCTTATCACTGATCCTTCCATCTTGTTCCTGGATGAGCCTACAAC 835

Qy      724 GGCCTCGACAGCTTTCACAGCCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGC 783
      | | | | | | | | | | | | | | | | | | | | | |
Db      836 GGCTTAGACTCAAGCACAGCAAATGCTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAG 895

Qy      784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
      | | | | | | | | | | | | | | | | | | | | | |
Db      896 GGACGAACAATCATCTTCTCCATTCATCAGCCTCGATATTCATCTTCAAGTTGTTTGAT 955

Qy      844 CTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTG 903
      | | | | | | | | | | | | | | | | | | | | | |
Db      956 AGCCTCACCTTATTGGCCTCAGGAAGACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTG 1015

Qy      904 CAGTACTTCACATCCATTGGCCACCCTTGTCTCTCGCTATAGCAACCCTGCGGACTTCTAC 963
      | | | | | | | | | | | | | | | | | | | | | |
Db      1016 GGATACTTTGAATCAGCTGGTTATCACTGTGAGGCCTATAATAACCCTGCAGACTTCTTC 1075

Qy      964 GTGGACTTGA 973
      | | | | | |
Db      1076 TTGGACATCA 1085

```

RESULT 13

US-09-961-086-2

; Sequence 2, Application US/09961086

; Publication No. US20030036645A1

; GENERAL INFORMATION:

; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE

; APPLICANT: ROSS, Douglas D.

; APPLICANT: DOYLE, L. Austin

; APPLICANT: ABRUZZO, Lynne

; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA

; TITLE OF INVENTION: WHICH ENCODES IT

; FILE REFERENCE: EP19376-019

; CURRENT APPLICATION NUMBER: US/09/961,086

; CURRENT FILING DATE: 2001-09-21

; PRIOR APPLICATION NUMBER: US 60/073,763

; PRIOR FILING DATE: 1998-02-05

; PRIOR APPLICATION NUMBER: PCT/US99/02577

; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-961-086-2

Query Match 6.6%; Score 132.4; DB 10; Length 2418;
Best Local Similarity 51.9%; Pred. No. 4.4e-30;
Matches 348; Conservative 0; Mismatches 316; Indels 6; Gaps 2;

```
Qy      304 GCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGA 363
        ||||| | || | |||| | || | | | || | || | || |
Db      467 GCCATCCTGGGACCCACAGGTGGAGGCAAATCTTCGTTATTAGATG---TCTTAGCTGCA 523

Qy      364 GGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCT 423
        | | | | | ||||| | || | ||||| || | | | |
Db      524 AGGAAAGATCCAAGTGGATTATCTGGAGATGTTCTGATAAATGGAGCACCGCGACCTGCC 583

Qy      424 CAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTG 483
        | | | | | || | | || | || | || | || | || |
Db      584 AATTTCAAATGTAATTCAGGT---TACGTGGTACAAGATGATGTTGTGATGGGCACTCTG 640

Qy      484 ACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCCAGGACCTTCTCCCAG 543
        || || |||| | | | || | || | || | || | || |
Db      641 ACGGTGAGAGAAAACCTACAGTTCTCAGCAGCTCTTCGGCTTGCAACAACCTATGACGAAT 700

Qy      544 GCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCC 603
        | ||||| | | || | || | || | || | || |
Db      701 CATGAAAAAACGAACGGATTAACAGGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCA 760

Qy      604 AACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTG 663
        || || | || | | || | | |||| | || | || | || |
Db      761 GACTCCAAGGTTGGAACCTCAGTTTATCCGTGGTGTGTCTGGAGGAGAAAGAAAAAGGACT 820

Qy      664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
        || || | || |||| | || | || | | ||||| || | || |
Db      821 AGTATAGGAATGGAGCTTATCACTGATCCTTCCATCTTGTTCTTGGATGAGCCTACAACT 880

Qy      724 GGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTGCTGGCCAAGGGC 783
        ||| | || | |||| | | | | | || | || | || |
Db      881 GGCTTAGACTCAAGCACAGCAAATGCTGTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAG 940

Qy      784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
        | | | |||| | || | ||||| | | ||||| | | |||| |
Db      941 GGACGAACAATCATCTTCTCCATTCATCAGCCTCGATATTCCATCTTCAAGTTGTTTGAT 1000

Qy      844 CTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTG 903
        || | | | || | || | || | || | || | || | || |
Db      1001 AGCCTCACCTTATTGGCCTCAGGAAGACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTG 1060

Qy      904 CAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTAC 963
        |||| | || | || | || | |||| | ||||| | |||| |
Db      1061 GGATACTTTGAATCAGCTGGTTATCACTGTGAGGCCTATAATAACCCTGCAGACTTCTTC 1120
```

Qy 964 GTGGACTTGA 973
| | | | | | |
Db 1121 TTGGACATCA 1130

RESULT 14

US-09-981-353-34

; Sequence 34, Application US/09981353

; Patent No. US20020160382A1

; GENERAL INFORMATION:

; APPLICANT: Lasek, Amy W.

; APPLICANT: Jones, David A.

; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER

; FILE REFERENCE: PA-0038 US

; CURRENT APPLICATION NUMBER: US/09/981,353

; CURRENT FILING DATE: 2001-10-11

; NUMBER OF SEQ ID NOS: 194

; SOFTWARE: PERL Program

; SEQ ID NO 34

; LENGTH: 2574

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CB1

US-09-981-353-34

Query Match 6.6%; Score 132.4; DB 9; Length 2574;
Best Local Similarity 51.9%; Pred. No. 4.5e-30;
Matches 348; Conservative 0; Mismatches 316; Indels 6; Gaps 2;

Qy 304 GCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGA 363
| | | | | | | | | | | | | | | | | | | | | |
Db 637 GCCATCCTGGGACCCACAGGTGGAGGCAAATCTTCGTATTAGATG---TCTTAGCTGCA 693

Qy 364 GGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCT 423
| | | | | | | | | | | | | | | | | | | | | |
Db 694 AGGAAAGATCCAAGTGGATTATCTGGAGATGTTCTGATAAATGGAGCACCGCGACCTGCC 753

Qy 424 CAGCTGGTGAGGAAGTGCGTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTG 483
| | | | | | | | | | | | | | | | | | | | | |
Db 754 AATTTCAAATGTAATTCAGGT---TACGTGGTACAAGATGATGTTGTGATGGGCACTCTG 810

Qy 484 ACCGTCAGAGAGACCCTGGCTTTTCATTGCCCAGATGCGCCTGCCAGGACCTTCTCCAG 543
| | | | | | | | | | | | | | | | | | | | | |
Db 811 ACGGTGAGAGAAAACCTTACAGTTCTCAGCAGCTCTTCGGCTTGCAACAACTATGACGAAT 870

Qy 544 GCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCC 603
| | | | | | | | | | | | | | | | | | | | | |
Db 871 CATGAAAAAACGAACGGATTAAACAGGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCA 930

Qy 604 AACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTG 663
| | | | | | | | | | | | | | | | | | | | | |
Db 931 GACTCCAAGGTTGGAACCTCAGTTTATCCGTGGTGTGTCTGGAGGAGAAAGAAAAGGACT 990

Qy 664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
| | | | | | | | | | | | | | | | | | | | | |

Db 991 AGTATAGGAATGGAGCTTATCACTGATCCTTCCATCTTGTCTTGGATGAGCCTACAAC 1050
 Qy 724 GGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCCGCCTGGCCAAGGGC 783
 ||| | ||| ||||| | | | | || | ||| |||
 Db 1051 GGCTTAGACTCAAGCACAGCAAATGCTGTCCTTTTGTCTGAAAGGATGTCTAAGCAG 1110
 Qy 784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
 | | || ||||| | || ||||| | | ||||| | | |||||
 Db 1111 GGACGAACAATCATCTTCTCCATTCATCAGCCTCGATATTCCATCTTCAAGTTGTTTGAT 1170
 Qy 844 CTGGTCCTTCTGATGACATCTGGCACCCCTATCTACCTGGGGGCGGCGCAGCAAATGGTG 903
 || | || | || | || | || | || | || | || | || | ||
 Db 1171 AGCCTCACCTTATTGGCCTCAGGAAGACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTG 1230
 Qy 904 CAGTACTTCACATCCATTGGCCACCCTTGTCTCGCTATAGCAACCCTGCGGACTTCTAC 963
 ||||| ||| ||| | | ||| ||||| ||||| ||||| |
 Db 1231 GGATACTTTGAATCAGCTGGTTATCACTGTGAGGCCTATAATAACCCTGCAGACTTCTTC 1290
 Qy 964 GTGGACTTGA 973
 ||||| | |
 Db 1291 TTGGACATCA 1300

RESULT 15

US-10-120-687-60

; Sequence 60, Application US/10120687

; Publication No. US20030082155A1

; GENERAL INFORMATION:

; APPLICANT: Massachusetts General Hospital

; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating Diabetes

; TITLE OF INVENTION: Mellitus

; FILE REFERENCE: 3284/1235B

; CURRENT APPLICATION NUMBER: US/10/120,687

; CURRENT FILING DATE: 2002-04-11

; PRIOR APPLICATION NUMBER: US60/169082

; PRIOR FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: US 09/963,875

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: US 60/215109

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: US 60/238880

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: US 09/731261

; PRIOR FILING DATE: 2000-12-06

; NUMBER OF SEQ ID NOS: 61

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 60

; LENGTH: 2718

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-120-687-60

Query Match 6.6%; Score 132.4; DB 14; Length 2718;

Best Local Similarity 51.9%; Pred. No. 4.7e-30;

Matches 348; Conservative 0; Mismatches 316; Indels 6; Gaps 2;

Qy 304 GCCATCATAGGGAGCTCAGGCTGCGGGAGAGCCTCACTACTCGACGTGATCACAGGCAGA 363
 ||||| | | | | | | | | | | | | | | | | | | |
 Db 433 GCCATCCTGGGACCCACAGGTGGAGGCAAATCTTCGTTATTAGATG---TCTTAGCTGCA 489

Qy 364 GGCCACGGTGGCAAGATGAAATCAGGACAAATTTGGATAAATGGGCAACCCAGTACGCCT 423
 | | | | | | | | | | | | | | | | | | | | | |
 Db 490 AGGAAAGATCCAAGTGGATTATCTGGAGATGTTCTGATAAATGGAGCACCGCGACCTGCC 549

Qy 424 CAGCTGGTGAGGAAGTGCCTTGCGCATGTGCGGCAGCATGACCAACTGCTGCCCAACCTG 483
 | | | | | | | | | | | | | | | | | | | | | |
 Db 550 AATTTCAAATGTAATTCAGGT---TACGTGGTACAAGATGATGTTGTGATGGGCACTCTG 606

Qy 484 ACCGTCAGAGAGACCCTGGCTTTCATTGCCAGATGCGCCTGCCCAGGACCTTCTCCAG 543
 | | | | | | | | | | | | | | | | | | | | | |
 Db 607 ACGGTGAGAGAAAACCTTACAGTTCTCAGCAGCTCTTCGGCTTGCAACAACCTATGACGAAT 666

Qy 544 GCCCAGCGTGACAAACGGGTGGAAGACGTAATCGCCGAGCTGCGGCTGCGGCAGTGCGCC 603
 | | | | | | | | | | | | | | | | | | | | | |
 Db 667 CATGAAAAAACGAACGGATTAACAGGGTCATTCAAGAGTTAGGTCTGGATAAAGTGGCA 726

Qy 604 AACACCAGAGTGGGCAACACGTATGTACGTGGGGTGTCCGGGGGTGAGCGCCGACGAGTG 663
 | | | | | | | | | | | | | | | | | | | | | |
 Db 727 GACTCCAAGGTTGGAACCTCAGTTTATCCGTGGTGTGTCTGGAGGAGAAAGAAAAGGACT 786

Qy 664 AGCATTGGGGTGCAGCTCCTGTGGAACCCAGGAATCCTCATTCTGGATGAACCCACTTCT 723
 | | | | | | | | | | | | | | | | | | | | | |
 Db 787 AGTATAGGAATGGAGCTTATCACTGATCCTTCCATCTTGTCTTGGATGAGCCTACAAC 846

Qy 724 GGCCTCGACAGCTTCACAGCCCACAATCTGGTGACAACCTTGTCCTCGCTGGCCAAGGGC 783
 | | | | | | | | | | | | | | | | | | | | | |
 Db 847 GGCTTAGACTCAAGCACAGCAAATGCTGTCTTTTGCTCCTGAAAAGGATGTCTAAGCAG 906

Qy 784 AACAGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGCTATTTGAC 843
 | | | | | | | | | | | | | | | | | | | | | |
 Db 907 GGACGAACAATCATCTTCTCCATTTCATCAGCCTCGATATTCATCTTCAAGTTGTTTGAT 966

Qy 844 CTGGTCCTTCTGATGACATCTGGCACCCTATCTACCTGGGGGCGGCGCAGCAAATGGTG 903
 | | | | | | | | | | | | | | | | | | | | | |
 Db 967 AGCCTCACCTTATTGGCCTCAGGAAGACTTATGTTCCACGGGCCTGCTCAGGAGGCCTTG 1026

Qy 904 CAGTACTTCACATCCATTGGCCACCCTTGTCCTCGCTATAGCAACCCTGCGGACTTCTAC 963
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1027 GGATACTTTGAATCAGCTGGTTATCACTGTGAGGCCTATAATAACCCTGCAGACTTCTTC 1086

Qy 964 GTGGACTTGA 973
 | | | | | | | |
 Db 1087 TTGGACATCA 1096

Search completed: February 27, 2004, 07:11:35
 Job time : 467.956 secs